Machine Learning and Monte Carlo based Data Analysis Methods in Cosmic Dust Research

A thesis accepted by the Faculty of Aerospace Engineering and Geodesy of the University of Stuttgart in partial fulfilment of the requirements for the degree of Doctor of Engineering Sciences (Dr-.Ing.)

by
Thomas Albin
born in Oldenburg (Oldb.), Germany

Main referee:	Priv.-Doz. Dr.-Ing. Ralf Srama
Co referee:	Prof. Dr. rer. nat. Alfred Krabbe
Co referee:	Prof. Dr. rer. nat. Eberhard Grün
Date of defence:	April 1st, 2019

Institute of Space Systems
University of Stuttgart
2019

Somewhere, something incredible is waiting to be known – Carl Sagan

Acronym and Abbreviations

AU Astronomical Unit 219

BIC Bayesian Information Criterion 6

CAT Chemical Analyzer Target 220

CCD Charged-Couple Device 196

CDA Cosmic-Dust-Analyzer 220

CILBO Canary Island Long-Baseline Observatory 216

DA Dust-Analyzer 103

DN Digital Number 184

DOCC Dust Orbit Computation Code 222

ECCS Earth-Centric Coordinate System 219

ECRC Earth-centric rotating coordinate system 218

FOV Field-of-view 220

GMM Gaussian Mixture Model 194

GP Gaussian Process 63

HCEC Helio-centric ecliptic coordinates 230

ICC Intensified CCD Camera 217

IDP Interplanetary Dust Population 228

IID Impact Ionisation Detector 220

IQR Inter Quartile Range 119

ISD Interstellar Direction 228

KDE Kernel Density Estimator 196

Inhaltsverzeichnis

Abbildungsverzeichnis

Listings

Vorwort

2014, während meines Masterstudiums an der Georg-August-Universität Göttingen und meiner Tätigkeit am Max-Planck-Institut für Sonnensystemforschung, verfasste ich meine Masterarbeit mit dem Titel "A global approach to analyse Cassini Cosmic Dust Analyser data using a relational data base system". Dr. Harald Krüger, der zum damaligen Zeitpunkt mein Betreuer und Mentor war, ermöglichte mir die Zusammenarbeit mit Dr. Ralf Srama. Aus der Arbeit heraus ergab sich dann das daten-getriebene Thema meiner Dissertation. Herr Srama hat mir viele Freiheiten gegeben und mich stetig sehr gut beraten. Ich konnte mich so schlussendlich zu dem Wissenschaftler und Data Scientist entwickeln, der ich jetzt bin. Ich hoffe, dass meine Arbeiten im Bereich Cosmic Dust und Cassini's Cosmic Dust Analyzer genug Daten, Impulse und Ideen geben, um weitere interessante Forschungsaufgaben anzuschließen. Ralf, das kurze Vorwort kann meinem Dank nicht gerecht werden, den ich hiermit aussprechen möchte.

Ein weiterer Dank geht auch an Herrn Prof. Dr. Alfred Krabbe und Herrn Prof. Dr. Eberhard Grün, die sich bereit erklärt haben meine Arbeit zu begutachten. Von beiden habe ich zahlreiche hilfreiche Kommentare erhalten und gute Diskussionen geführt. Dank Ihnen, wurde ich zum Ende hin nicht allzu "betriebsblind".

Herrn Prof. Dr. Björn Poppe und Herrn Dr. Detlef Koschny gebührt mein Dank bei der Zusammenarbeit am Meteorsystem Canary Islands Long-Baseline Observatory. Herr Koschny war schon zu meinen Bachelorzeiten ein großartiger Mentor, mit dem ich gerne zusammengearbeitet habe. In diesem Rahmen geht auch mein Dank an Herrn Dr. Gerhard Drolshagen. Bei Frau Prof. Dr. Jutta Kunz-Drolshagen bedanke ich mich für die Unterstützung und Förderung meines astro-physikalischen Interesses.

Ich danke auch all denen, mit denen das (persönliche) Leben am IRS Spaß gemacht hat. Hierbei bedanke ich mich vor allem an meine Kollegen aus der Staubgruppe, allen voran Jonas Simolka und Rachel Halina Soja. Bei Bartomeu Massuti-Ballester bedanke ich mich für die unzähligen gelaufenen Kilometer um den Bärensee. Auch dank der Spielegruppe um Manfred Ehresmann und Jonathan Skalden hat mir die Zeit nach der Arbeit viel Freude bereitet. Zu guter Letzt geht ein besonderer Dank an Adam Pagan und die aus gemeinsamen Interessen ergebene Freundschaft.

Die Arbeit wäre auch nicht möglich gewesen ohne die Zusammenarbeit mit vielen Studenten, die bei mir ihre Abschlussarbeiten verfasst haben und teilweise als studentische Hilfskraft weitergearbeitet haben. Ich bedanke mich hier bei Max Faber, Lena Bötsch, Felix Burkhardt und allen voran Bastian Knieling und Felix Willenborg.

Mit dem Schlusswort bedanke ich mich an die “Oldenburger Gruppe” um Gerd Liebig, Hannah Ramsauer, Sarinah Sutojo und Marcel Kahlen. Danke auch an Marius Hacker. Ein großer Dank geht auch an meine besten Freunde Henning Elbers und Holger Diedrichs, die mich seit der Schule begleiten.

Meine Eltern Agnieszka und Marek Albin gebührt ein ganz besonderer Dank. Sie unterstützten mich in allen Phasen meines Lebens, ermutigten und berieten mich und haben meine Interessen stets gefördert.

Zuletzt geht ein großes Dankeschön an meine Freundin Stefanie Arend. Mit dem Abschluss meiner Promotion beginnt ein neuer Lebensabschnitt mit Dir auf den ich mich mit diesem Schlusswort dieser Arbeit sehr freue.

Kurzfassung

Künstliche Intelligenz, Big Data oder das Petabyte-Zeitalter sind Schlagwörter, die die IT Welt zu Beginn des 21. Jahrhunderts prägen. Ob im Bereich Social Media, Industrie 4.0 oder bei den Internet of Things; moderne Algorithmen dienen dazu Prozesse zu automatisieren, Benutzer zu analysieren oder ihnen eine intelligente Benutzung von Geräten zu ermöglichen. Zusätzlich werden in zahlreichen Bereichen numerische Methoden verwendet, um nicht-analytische Problemstellungen statistisch zu beschreiben. Und mit immer fortschrittlicherer Hardware ist es schon heute möglich, neuronale Netze und andere rechenintensive Prozesse auf mobilen Endgeräten zur Verfügung zu stellen.

Diese Arbeit wendet ausgewälte, verfügbare Algorithmen auf den Wissenschaftsbereich *Cosmic Dust* an, um das Potential für diesen Forschungszweig auszuloten. Im Rahmen der Arbeit werden technische und wissenschaftliche Daten aus zwei Projekten verwendet: Das stereoskopische Meteorbeobachtungssystem Canary Island Long-Baseline Observatory (CILBO), welches aus zwei autonomen Observatorien auf Teneriffa und La Palma besteht; sowie der Cosmic-Dust-Analyzer (CDA), welches ein Staubteleskop auf der Sonde Cassini war, um in-situ kosmische Staubpartikel im Saturnsystem zu detektieren. Zu Beginn dieser Arbeit erfolgte der wissenschaftliche Betrieb von CILBO und CDA, welcher dann im Laufe der Zeit durch technische Defekte seitens der Meteorkamera und den Sturz der Cassini Sonde in den Saturn unterbrochen, bzw. beendet worden ist.

Sowohl für das CILBO, als auch CDA Projekt sind Monte Carlo basierte Programme zur Berechnung astrodynamischer Eigenschaften von Staubteilchen entwickelt worden. Hierbei sind die astrometrischen Ungenauigkeiten bei der beobachtenden Astronomie, sowie das große Gesichtsfeld und der Fehler bei der Bestimmung der Geschwindigkeit bei in-situ Instrumenten die a-priori Wahrscheinlichkeitsfunktionen der Programme. Vor allem für das CDA Projekt sind Kernel Density Estimator (KDE)s verwendet worden, um diese a-priori Funktionen zu bestimmen.

Mit Hilfe der CILBO Orbit Datenbank lassen sich so in Zukunft wissenschaftliche Analysen zu Meteorströmen und -sporaden durchführen. Ein statistischer Überblick über die Daten zeigt vor allem den Umfang der registrierten Sporaden aus Apex- und Antihelion-Richtung. Des weiteren erfolgt eine technische Bias Analyse von CILBO. Mittels KDEs und anderen Algorithmen zur Identifikation von Datenstrukturen werden diverse Bias-Effekte beschrieben, die beispielsweise auf die Beobachtungsgeometrie beider Stationen zurückzuführen sind.

Auch für CDA entstand so eine umfangreiche Datenbank mit hunderttausenden Einträgen.

Mit Hilfe der Monte-Carlo (MC) generierten Ergebnisse zur Teilchendynamik erfolgt in der Arbeit eine Neubestimmung der Upstream-Richtung der interstellaren Staubteilchen. Ein Vergleich mit den Daten vergangener Publikationen zeigt, dass sich die Richtung in den letzten Jahren nicht signifikant verschoben hat.

E-Ring Teilchen in der CDA Orbit Datenbank zeigen einen Bias zu hohen Geschwindigkeiten: Die meisten Teilchen wurden laut der MC Methode im Geschwindigkeitsmaximum, also in der Periapsis gemessen. Dieser Bias zu vermehrten Messungen in der Periapsis von Staubteilchen resultiert in der Annahme eines Geschwindigkeitsbestimmungsbiases. Neuronale Netze werden anschließend verwendet, mit denen 5 neue Geschwindigkeitskalibrationen von CDA erfolgen. Unter anderem wird so erstmals eine Funktion für Wandeinschläge bestimmt. Anstiegszeit- und ampltiudenabhängige Funktionen werden abgeleitet, die im Median geringere Geschwindigkeitswerte bestimmen als die ursprünglich verwendeten Kalibrationsfunktionen.

Anschließend erfolgt eine Typ-Klassifizierung der Staubteilchen. In Kooperation mit den Studenten B. Knieling und F. Burkhardt sind supervised Machine Learning Algorithmen entwickelt, trainiert und auf unbekannte Daten angewandt worden, um zwischen Typ 4 und Nicht-Typ 4 zu differenzieren. Dieser Mineralspektrentyp wird beispielsweise mit interstellarem Staub assoziiert.

Insgesamt versteht sich diese Arbeit als Kompilation verschiedener MC und Machine Learning (ML) basierter Projekte, angewendet auf CILBO und CDA. Clustering, KDE, supervised classification, unbiased statistics, etc. decken diverse Themenfelder ab und sollen zugleich die Basis für weitere wissenschaftliche und technische Fragestellungen sein. Vereinfacht wird die mögliche Fortsetzung der Arbeiten mit Hilfe verschiedener Anwendungen, die in PYTHON entwickelt und nach PEP8 Coding Standard geschrieben und dokumentiert worden sind.

Abstract

Artificial Intelligence, Big Data or the Petabyte-Age are current IT buzzwords of the starting 21st century. Modern algorithms automatise the analysis of users and the content in social media, control and link the hardware of industry 4.0 products or the Internet of Things. Recent hardware developments allow also the implementation of sophisticated processes like e.g. neural networks on mobile devices. Thus the usage of complex devices becomes more comfortable by using voice- or gesture-commands. Additionally, with more computational power, numerical calculations can be performed to describe non analytical problems in a statistical way.

This work applies miscellaneous algorithms from the fields *Machine Learning* and *Computational Numerics* on the research field *Cosmic Dust*. The task is to determine the scientific and technical potential of using different methods. Here, the methods are applied on two different projects: the meteor camera system Canary Island Long-Baseline Observatory (CILBO) and the Cassini in-situ dust telescope Cosmic-Dust-Analyzer (CDA). CILBO is a stereoscopic system, that is installed on La Palma and Tenerife, respectively. During this thesis both systems operated perfectly and delivered a large amount of data. However, possibly due to material fatigue and due to Cassini's mission ending, the operation of CILBO and CDA ended in the last couple of months.

For both instruments a Monte-Carlo based orbit computation tool has been developed, to determine several dynamical properties of meteors and in-situ dust particles, respectively. For the CILBO programme (that is currently embedded into the Meteor Research Group's Meteor Orbit and Trajectory Software (MOTS) toolkit), the astrometric positional accuracy and derived velocity uncertainties are the input Probability Density Function (PDF)s for the code. On the other hand, a few Kernel Density Estimator (KDE) approaches have been used for CDA to derive proper a-priori functions (e.g., the angular Field-of-view (FOV) of both targets needed to be modelled to consider the directional uncertainties). Altogether, two vast databases have been generated, that allow further scientific and technical analysis. For example, with the CDA database known possible dust candidates of interstellar origin have been dynamically analysed. A new upstream direction, that is a few degrees off the radiant determined by previous research has been computed, too.

Furthermore, the CDA database reveals a possible velocity bias. The analysis of Saturn's E-Ring particles shows, that most particles have been detected in the vicinity of their corresponding periapsis, reaching their maximum velocity. Thus, a neural network based approach has been performed to re-calibrate the instrument's velocity function. For each target 2 new

functions have been calculated and further, a wall impact function has been derived too. The resulting velocities got smaller (w.r.t. to the originally used functions) and will be implemented in the future in new dynamical studies.

Another detailed technical bias analysis applies on the CILBO data. Different observational geometries and time depending radiant changes of the sporadic sources show, that both systems detect meteors with different velocity and brightness distributions. Although both systems are similarly build, miscellaneous effects lead to a shift in the overall distributions. Additionally, instrument optics, detection algorithms etc. lead to further uncertainties, resulting in a multi-dimensional bias problem, that needs to be considered for flux and velocity computations.

In cooperation with the two students B. Knieling and F. Burkhardt supervised machine learning classification tools have been developed to identify yet unknown type 4 spectra in the CDA database. These so called mineral spectra are associated with dust from interstellar space or from yet unknown parent bodies or Saturnian satellites. Based on a few hundred and already labelled spectra, the set of new type 4 data raises by a factor of around 2.5 and is now a basis for further scientific research.

Altogether this work is a compilation of miscellaneous approaches to analyse scientific and technical data from cosmic dust research instruments. Since two different measurement principles have been considered, namely a visual observational one and an in-situ detector, several different aspects have been covered. Different approaches lead to an unbiasing of the data, re-computation of calibration functions and new numerical calculations. The techniques are a sophisticated basis for future analysis and the two databases cover enough information for several research projects.

Additionally, most developed PYTHON scripts have been written and documented based on the PEP008 coding standard. This will allow further and quicker developments in the future.

1. Introduction

The scope of this work is to combine modern and recent developments in data analysis with space science applications. Hereby, the focus is set on cosmic dust instrumentations, namely the ground-based, stereoscopic video camera system Canary Island Long-Baseline Observatory (CILBO) and Cassini's on-board instrument Cosmic-Dust-Analyzer (CDA) that detected dust particles in-situ during an interplanetary cruise-phase and within the Saturnian system.

The thesis is considered as a comprehensive compilation of miscellaneous Machine Learning (ML), data scientific and numerical, Monte-Carlo (MC) based projects. Since CILBO and CDA are separate scientific projects, the chapters are not linked. The reader can choose freely the interesting chapters and sections, respectively. This work applies miscellaneous approaches for different scientific and technical questions and topics. The results, generated databases and used methods are a solid basis for further projects and theses and shall support future students to continue the set basis in different disciplines. E.g., the developed Neural Network (NN)s for the CDA velocity calibration (5.4) can be used for other calibration purposes; or the developed orbit databases (CILBO and CDA), that contain a vast set of data points for scientific projects (e.g., Interstellar Direction (ISD) particles identification on Earth, E-Ring analysis, Phoebe-Ring analysis, Rhea dust-cloud formation or the analysis of dust particles emerging from the Tiger-stripes on Enceladus).

Detailed overviews of the data scientific methods and the instruments are given in the corresponding chapters.

1.1. Structure of this thesis

This thesis consists of 4 chapters. The first two chapters introduce statistical and data scientific methods and an astrodynamical background, respectively. Afterwards two chapters are about CILBO and CDA.

First, a few basics on ML and MC simulations are shown. Both fields are vast and cannot be shown in detail. Only the theoretical background of concepts is shown, that is used later for actual instrument data analysis. Furthermore artificially generated datasets provide application-oriented examples to visualise miscellaneous algorithms.

The second theory chapter gives an overview on astrodynamics. The Keplerian elements are explained and formulae are provided for computational purposes. Since several algorithms that have been developed during this thesis have several SPICE implementations, reference

frames are shown and the corresponding kernels are listed in the appendix. A section provides information on gravitational sphere concepts like the Sphere Of Influence (SOI) and Hill sphere. Further, the zenith attraction, D-Criterion and Tisserand parameters are explained as well as radiant information that are mostly used in meteor research.

Afterwards a chapter provides detailed insights of the CILBO system. For the CILBO part two projects have been performed. A technical bias analysis uses i.a. Kernel Density Estimator (KDE) methods to analyze the recorded data of the meteor camera system to understand and unbias the obtained data sets [Albin et al., 2017]. Conclusions of this work are then used in the second project, a MC based orbit computation code, to properly determine the astrodynamical properties of stereoscopically recorded meteors. The analysis and the new established orbit database are a basis for further technical and scientific projects.

The last major chapter focuses on Cassini's CDA. The MC based Dust Orbit Computation Code (DOCC) is shown, that provides a large astrodynamical database for further scientific purposes. One major scientific project is the dynamical analysis of the ISD particle candidates by Altobelli et al. [2016] and is shown afterwards. Since the number of exogenic particles appears to be too large in the database, a velocity calibration bias is assumed. Consequently this leads to a project to determine new velocity functions for the different instrument's channels. Srama [2000] used polynomial function to fit the calibration data. Here, a new NN based approach is performed to determine a multi-dimensional regression plane. Finally, a ML based spectra classification is shown that focuses on the search of type 4 spectra. These type 4 spectra correspond to mineral dust particles with no water-ice and are associated with retrograde particles from Phoebe or exogenic particles from interplanetary or interstellar space [Altobelli et al., 2016].

2. Advanced Data Analysis Methods

This chapter describes some basic concepts in Machine Learning and Monte-Carlo theory. Both concepts are used in this work to evaluate and perform modern data analysis driven methods on Earth-bound and in-situ dust observations. The methods are later used to determine technical biases, to perform new calibration approaches, or to compute and generate derived data output and databases for future scientific purposes.

The historic background, basic mathematical concepts, sub-topics and applications in Machine Learning and Monte-Carlo fields, are extensive and complex. Due to the recent increase of computation performance, cloud computing and the growing amount of Big Data, this chapter provides only an overview. Those methods are shown, which are later used. Furthermore, a few examples simplify theoretical ideas to allow a more comprehensible access for space scientists, who are not working in this particular field.

Especially the following references provide more details and a sophisticated overview on certain concepts. Methods, Algorithms and concepts are partly adapted, modified and extended in this work, based on the provided information given by the books from:

Bishop [2011] provides a definitive book for machine learning. The extensive content gives a detailed mathematical overview of miscellaneous machine learning concepts.

Ivezić et al. [2014] provide a sophisticated overview and introduction in several Machine Learning (ML) fields that fulfils several demands. The book describes algorithms and methods qualitatively as well as mathematically, but not as extensive as Bishop [2011]. Furthermore, several real-data examples are provided with code snippets to reproduce the shown methods. The authors focus on applying the methods on astrophysical data sets, however, astrophysical knowledge is not required though. Based on the book, the PYTHON library ASTROML has been developed (see below).

Cristianini and Shawe-Taylor [2000] provide a detailed overview of the particular ML method: Support Vector Machine (SVM).

Furthermore, the following software packages have been used, with their according documentation that further improve the usage and application of ML algorithms:

scikit-learn

This library [Pedregosa et al., 2011] is the most commonly used package for Machine Learning applications. It contains several algorithms for clustering, regression and classification problems. Preprocessing, feature extraction and dimension reduction methods are implemented, too.

astroML

Developed by Vanderplas et al. [2012] and Ivezić et al. [2014], this library uses SCIKIT-LEARN modules for astrophysical applications. Several extensions of rather exotic and recent methods, like e.g. extreme deconvolution [Bovy et al., 2011], are implemented.

TensorFlow

The library is currently the most sophisticated solution to develop, implement and use neural networks [Abadi et al., 2015]. A native graphics processing unit support (using NVIDIA's CUDA library) allows parallelised computations. Developed by the GOOGLE BRAIN TEAM, this library is used in several GOOGLE applications.

Keras

Chollet et al. [2015] introduced a framework package for two neural network packages. One supported library is TENSORFLOW. The library simplifies TENSORFLOW's programming structure and allows a faster and more agile way of developing Neural Networks using the most common features.

2.1. Machine Learning

ML is a general term for self-learning or adaptive algorithms, allowing a computer program "to learn" patterns, data structures, features etc. First ideas were proposed by Rosenblatt [1958], who developed the first perceptron algorithm, based on miscellaneous previous works in the 1940s. The vast field can be divided in two general approaches: supervised and unsupervised learning. Both approaches have several sub-fields. Supervised approaches are based on a well known data set that is labelled, e.g., mass spectra with different classes. An algorithm adapts then to this well known data set. Further, via cross-validation, the classification precision (score) can be determined. Unsupervised approaches do not use any labelled data and try to adapt automatically on the data. However, miscellaneous metrics are available that determine quantitatively the "quality" of the resulting model. Further, these algorithms allow a model tweaking with so-called hyper-parameters that adjust certain settings or alter the modelling behaviour. Examples on this topic are in the following sections. A third field does not fit perfectly in either approaches: Reinforcement Learning. This method is in a way unsupervised, but the algorithm receives a positive or negative signal, depending on its behaviour and decision. Example: A self-learning algorithm to play a video game, where the positive or negative signal is e.g., a won or lost game. The mechanics of the game are learned unsupervised without any user given information (meaning and purpose of keys). This particular field is not used in this work.

In the following 5 application areas Machine Learning algorithms are currently used:

Regression

Multidimensional, non linear regression models that mix continuous with discrete data.

Classification

Binary, or multi-classified data that are learned by a model and later applied on an unclassified data set.

Clustering

Identification and description of data clusters in multi-dimensional space.

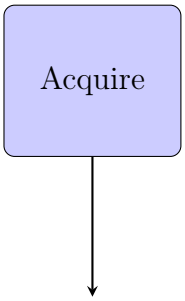
Density Estimator

Modelling the density distribution of (discrete) multi-dimensional data points.

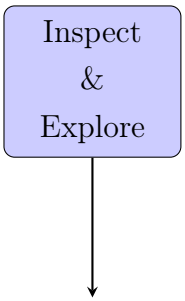
Dimension Reduction

Reducing a high dimensional data set and describing it by lower-dimensional properties.

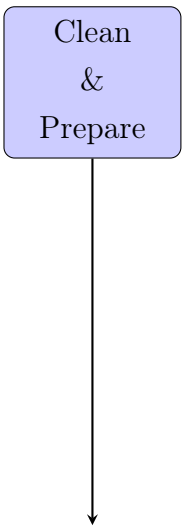
To develop a proper ML model based on a data set, the following framework is recommended [Combs, 2016]. The framework is based on 6 development steps, from the data acquisition to the final deployment (and later implementation). Especially the preprocessing, cleaning of the data and the well set requirements for the model are important steps to choose a proper ML algorithm.



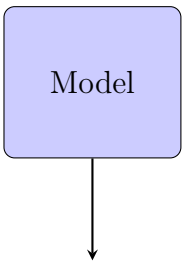
The very first step in Machine Learning is the data acquisition. This depends on the scientific or technical area and can be divided generally in two data fields: static and streaming data. Static data are, e.g., obtained from the instruments shown in this work. A certain event is recorded and stored. Afterwards, the data do not change. Streaming data are, e.g., continuous audio or video streams that need to be analysed in real-time.



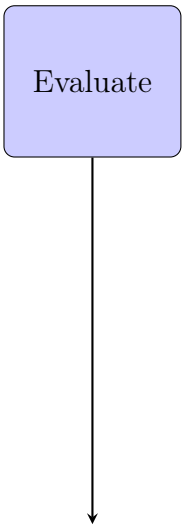
The inspection and exploration phase is done mostly manually by a user. Consistency checks are performed, possible wrongly determined data are sort out (e.g., checksum errors) and all necessary data are then transformed into a needed or proper format. Furthermore, the user defines the tasks, an algorithm needs to fulfil and sets the according requirements to determine the corresponding ML approach.



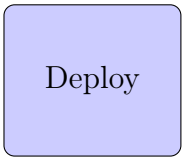
This phase is used to prepare the data for the according algorithms. The preparation of the data contains two sub tasks: first, the data need to be "normalized" or "standardized", respectively. This means that certain statistical properties need to be be adjusted for the algorithms. E.g., most SCIKIT-LEARN functions require a mean value of 0 and a standard deviation of 1 to work properly. Depending on the dimensionality of the data set and the task, feature extractions need to be applied on the data to improve the performance of the algorithms (for example Fourier-Transformation of a signal and extracting the relevant frequencies). Further details can be found here `http://scikit-learn.org/stable/data_transforms.html` (01.03.2018).



The prepared data set can now be used to generate a Machine Learning model, respectively train an algorithm. The chosen model depends on the requirements, whether it is a regression, clustering or classification problem; and whether it is supervised or unsupervised. This step is linked with the following step...



...the evaluation phase. Several model hyper-parameters need to be adjusted or iteratively tested to determine the ideal model. Since this is mostly an optimization problem that cannot be determined analytically, e.g., within a for-loop, several parameters are tested. Also, the evaluation depends on the chosen algorithm. Supervised learning methods with labelled data are separated in a training and validation sub-set, where the validation set is used to evaluate the model e.g. with a confusion matrix (see later in the subsection regarding classifications 2.1.2). For unsupervised methods with hyper-parameter adjustment miscellaneous metrics are available, like the Bayesian Information Criterion (BIC) that will be introduced later.



Based on the optimal model parameters the final Machine Learning model is generated. This model is then used and applied on further, unknown / unlabelled data.

This work uses the first 4 algorithm approaches. Especially chapter 4 uses miscellaneous *Density Estimators* to identify technical bias effects. *Regression*, *Classification* and *Clustering* algorithms have been applied in chapter 5, where ML attempts have been performed to re-

calibrate the dust instrument or to analyse scientific data. Further, Albin [2014], Burkhardt [2017] and Knieling [2017] used *Classification* algorithms to classify certain mass spectra in the Cosmic-Dust-Analyzer (CDA) database.

The following sub-sections introduce and describe selected methods and algorithms, partly based on examples. Due to the vast number of algorithms, Babcock [2016] provides a few summarizing flow charts that support users to determine a problem-depending, suitable algorithm. On p. 81, he provides a chart to determine a proper distance metric / method. P. 100, 141 and 181 provide a decision chart to choose a proper algorithm for clustering, regression and classification problems, respectively.

2.1.1. Structures in data sets

To determine underlying structures, distributions or descriptive parameters of a (multi-dimensional) data set that was, e.g., generated by a simulation or recorded from an experiment, miscellaneous methods are available.

Most researchers apply by default Gaussian measures, like the mean or standard deviation. However, this method is only suitable for datasets that are Gaussian distributed or at least symmetric. For 1- or 2-dimensional datasets a Gaussian or symmetric structure can be determined e.g. by visual inspection. Higher dimensional data require more sophisticated solutions to determine its symmetry, or, one can apply more robust measures on the data set.

Robust measures Consider the following distribution, shown in figure 2.1. The Y axis shows a (density) distribution of an artificially generated distribution, thus, the X axis is the variable space (e.g., measurement points). This example has arbitrary units. The generated data points are plotted at the bottom of the plot and the black solid curve shows the corresponding distribution that generated these data, however, the distribution and parameters are unknown for the analyst. Further, the distribution is not symmetric or Gaussian. Most commonly, a histogram is a first data visualisation attempt to determine the structure of the obtained data points (light blue histogram in figure 2.1). The number of bins g that include a set of data points should not be chosen randomly. Several methods are available to set this parameter. A commonly used "rule-of-thumb" is, where n is the number of data points, embedded in the floor function:

$$g = \sqrt{\lfloor n \rfloor}$$

The resulting histogram does not describe the actual distribution smoothly. It describes the descending character for higher variables, but does not visualize the peak of the function.

More robust attempts were introduced by Tukey [1977], who defined the explorative data analysis in the mid 20^{th} century. Above the histogram is a horizontally aligned box plot shown

that contains the following information: A blue line within a drawn box indicates the median of the distribution. The box itself defines the Inter Quartile Range (IQR) around the median and contains 50 % of the data. The horizontal, solid lines with endings that are attached to the box (left and right side) are the whiskers and have, as defined by Tukey [1977], a maximum length of 1.5 times the IQR. Other definitions set the whiskers at 2.5 % and 97.5 % of the data; consequently the whiskers contain 95 % (confidence interval) of the data. With Tukey's definition, the following rough rules apply for data that are not within the whiskers (in the figure indicated as triangles): Data that lay between 1.5 and 3.0 times the IQR are so called "mild" outliers, while larger distances indicate a "severe" outlier.

The box plot shows also two indentations around the median value. These so called notches [Chambers et al., 1983] can be used to determine whether the median values of two box plots are significantly different or not, by determining whether the corresponding notches overlap or not, respectively. The width w_{notch} of the notches is computed as:

$$w_{\text{notch}} = 1.57 \cdot \frac{\text{IQR}}{\sqrt{n}}$$

An additional robust measure is the so called Median Absolute Deviation (MAD) that indicates the median deviation between the data points x_i (with i as the index) and the median of these points:

$$\text{MAD} = \text{median}\left(|x_i - \text{median}\left(x_i\right)|\right)$$

Assuming a symmetric or Gaussian distribution, the standard deviation σ is linked with the MAD the following way: $\sigma = 1.4826 \cdot \text{MAD}$.

In figure 2.1 the median and corresponding MAD values are shown above the box plot. Due to the larger number of outliers and the skewness of the function, the computed mean and standard deviation (in brackets, shown above the median value) do not correspond to the median and scaled MAD. The mean is shifted towards larger values and also the standard deviation is in the same order as the mean itself. Both values do not describe the underlining function properly. A red dashed line indicates the position of the mean within the box plot.

The box plot, with all its determined features and the MAD are used in this work especially in determining statistical properties of Monte-Carlo results, e.g. for the orbit determination software for the CDA (see section 2.2).

Kernel Density Estimator (KDE) As shown, histograms cannot determine finer structures of a data set. Overfitting or under-fitting are the most common issues that lead to undiscovered structures or over-interpretation. Robust measures like box-plots can be used to determine statistical descriptive parameters of a function. However, the skewness or other important

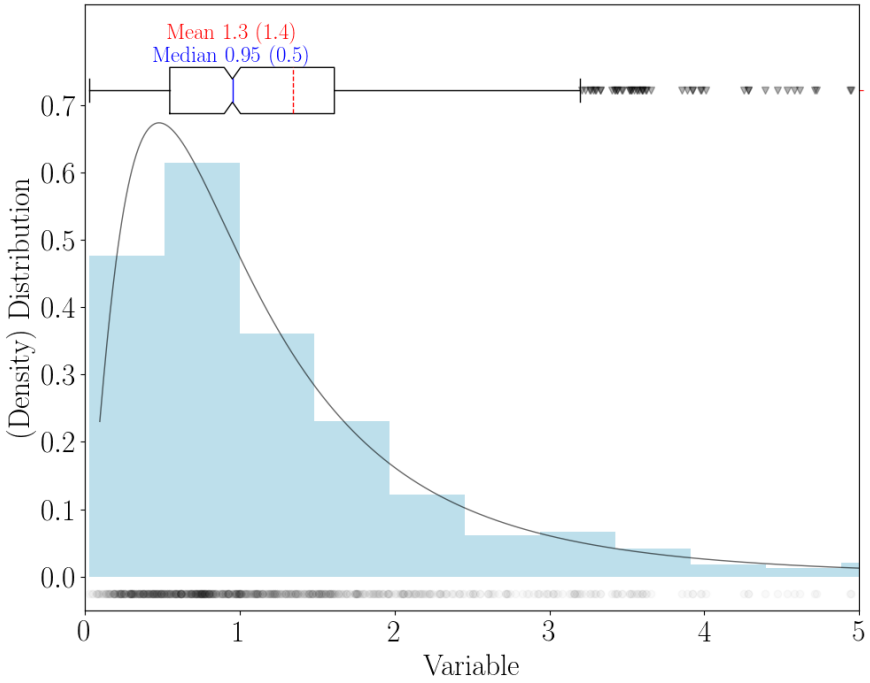


Figure 2.1.: Visualization of an artificially generated data set. The density distribution is plotted vs. the set variable. A black solid curve indicates the underlying distribution that generates the data points, shown in the bottom of the graph. A histogram shows the draw-backs of this method: The maximum of the distribution cannot be display properly. Above is a horizontally aligned box plot shown. A dashed line indicates also the mean of the distribution. Both values, median and mean, with corresponding Median Absolute Deviation (MAD), standard deviation, are displayed in blue and red, respectively. The box itself contains 50 % of all data points and the symmetric indenations around the median are the so called notches. These notches are used to determine if the median of two data sets are significantly different or not. So called whiskers are attached to the box plot. Depending on the definition, the wiskers have a maximum of length of 1.5 times the interquartile range or include a certain confidence interval. Data points that are not included by the wiskers are shown as triangles and indicate outliers.

features cannot be clarified. Further, this method might not be suitable for distributions with several peaks or features.

To determine the structure of (multi-dimensional) data sets, without any a-priori knowledge regarding the underlying structure, the KDE method can be applied.

KDE algorithms replace a given data point x_i with a function K. This function, also called *Kernel*, "smears" the discrete data points over a certain domain. The superposition of all Kernels generate the final, continuous distribution. The final distribution $\hat{f}_{n,h,D}(x)$ is determined by the following function [Ivezić et al., 2014, Sec. 6.1.1]: D is the number of dimensions. K is the Kernel with a certain bandwidth h. The x variable depending value $\hat{f}_{n,h,D}(x)$ is the sum of all kernels K, where the weight or functional contribution of each Kernel depends on the distance between x and each Kernel centre, defined by x_i.

$$\hat{f}_{n,h,D}(x) = \frac{1}{n \cdot h^D} \cdot \sum_i^n K\left(\frac{d\,(x, x_i)}{h}\right)$$

Kernels have to fulfil miscellaneous properties, like e.g. returning only positive semi-definite values at any position. Most common kernels are the Epanechnikov, Exponential, Tophat, or Gaussian Kernel.

The bandwidth, so the definition range or length of a kernel (standard deviation for the Gaussian Kernel) is the only Kernel-hyper-parameter that needs to be set accordingly. The mathematical correct way to determine the bandwidth is to compute the *Mean Integrated Squared Error* of $\hat{f}_{n,h,D}(x)$ and the function that is the actual distribution function. Since this method requires the, in most instances, unknown distribution function, several computational methods have been developed. One commonly used algorithm is the *Cross Validation* that is described in Ivezić et al. [2014, p. 254] and implemented in several data scientific libraries like SCIKIT-LEARN[1]. This method computes the normalised sum of *one-leave-out* kernel functions (using all n data points, but the i^{th} one, indicated with a minus). The resulting set of cross validation functions CV needs than to be optimised with respect to the bandwidth h.

$$CV(h) = \frac{1}{N} \sum_{i}^{n} \log \hat{f}_{-i,h,D}(x)$$

However, several "rules-of-thumb" can be applied for a quick overview on the dataset. For example Venables and Ripley [2003, p. 127/128] lists the following two rules for a Gaussian Kernel. Both bandwidths describe an upper boundary.

$$h_{\text{gauss},1} = 1.06 \min\left(\sigma,\ \text{IQR}/1.34\right) \cdot n^{-1/5} \tag{2.1}$$

$$h_{\text{gauss},2} = 1.144 \cdot \sigma \cdot n^{-1/5} \tag{2.2}$$

Scott [2015, Tab. 6.3] provides a list with different coefficients to convert the bandwidth determined by equation 2.2 to bandwidths for different kernels. Equation 2.3 shows the "rule-of-thumb" bandwidth for the *Epanechnikov* kernel.

$$h_{\text{epan.}} = 2.214 \cdot h_{\text{gauss},2} \tag{2.3}$$

More sophisticated approaches determine local density fluctuations and set the bandwidth adaptively for each data point.

An example of an applied KDE can be seen in figure 2.2. The figure shows an artificially generated data set, based on two Gaussians with a mean and standard deviation of $\mu_1 = -1;\ \sigma_1 = 1$ and $\mu_2 = 5;\ \sigma_2 = 2$, respectively, and an according weighting of 0.7 and 0.3.

[1] Further implementation details: `http://scikit-learn.org/stable/modules/cross_validation.html`, 19.03.2018

The $n = 400$ generated data points are visualized in a histogram and a red-dashed curve indicates the actual distribution, based on the Gaussian properties. The black solid curve shows a Gaussian kernel based KDE. The KDE describes the according peaks sufficiently, but the chosen bandwidth appears to be too broad. Since $h_{\text{gauss},2}$ has been used in this figure the according bandwidth was set at its upper possible boundary.

Above the KDE is a so-called violin-plot drawn. This graph is basically a KDE, that is plotted symmetrically, recalling the shape of a violin instrument. In combination with a box-plot an analyst can determine whether a box-plot provides sufficient results. A black cross shows the median.

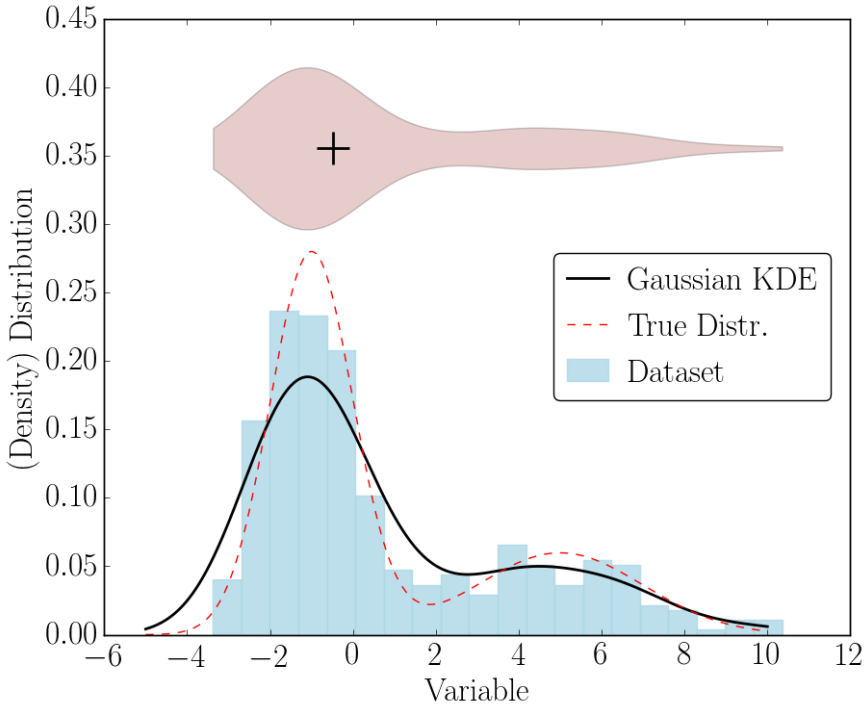


Figure 2.2.: Miscellaneous visualization methods of an artificially generated data set. The density distribution is plotted vs. the set variable. A red dashed curve shows the actual distribution and the black solid curve represents the Gaussian-kernel based Kernel Density Estimator (KDE) approach to determine the underlying distribution. A histogram is shown, too. Above the distribution is a violin-plot visualized with a black cross that indicates the median of the data set. A violion-plot is symmetrically drawn Kernel Density Estimator (KDE) plot.

Gaussian Mixture Model (GMM) KDE algorithms determine the overall structure of a multi-dimensional data set without any a-priori knowledge. In some cases, a-priori knowledge is available for the analysis of data sets. The normalised histogram in figure 2.3 shows an artificially generated data set, based on 3,000 data points. The corresponding distributions are 3 Gaussians, with a certain mean μ and standard deviation σ listed in the second column in table 2.1. The distribution shows a major peak and a skewness, indicating that the data cannot be described by one single Gaussian.

For example, a-priori, an analyst might know the following two properties: first, the data can be described by Gaussian functions; second and optionally, 3 Gaussians are responsible for the

generated data. In this case a GMM algorithm can be applied that "tries" to fit the number of user-set Gaussians on the data. The major hyper-parameter is the number of distributions. The properties like mean and standard deviation are determined automatically with miscellaneous methods. For higher dimensional data sets, general covariance matrix attributs can be set, too.

In this case a for-loop iterates through a number of Gaussians, from 1 Gaussian fit to 10. For each number of Gaussians (hyper-parameter k) the so called BIC was computed. The BIC is defined as, where $L_{\max}$ is the maximum value of the model's likelihood function:

$$\mathrm{BIC}(k, L_{\max}) = \ln(n)k - 2 \cdot \ln(L_{\max})$$

Qualitatively speaking, the BIC is a mathematical description of *Occam's Razor*: The principle that the "simplest" solution with least descriptive parameters and hypothesis is the preferred solution for a problem.

The BIC value increases for a larger number of parameters k (here: number of chosen Gaussians), however decreases on the other hand, if the corresponding likelihood function overcompensates the number of parameters. Thus, using the BIC to determine the number of solutions is an optimisation problem that needs to be minimised. Models with a better likelihood function get penalised by the larger number of chosen parameters. Figure 2.4 shows the BIC value vs. the number of chosen Gaussians. For better visibility, the dots are connected with solid lines. The information criterion decreases for more Gaussians and reaches a minimum at $k = 3$ (indicated with a red circle). Afterwards, the BIC increases again, with slight fluctuations. Thus, a higher number of Gaussians does not improve the likelihood significantly and the larger number of ks affect the BIC negatively.

The found Gaussians are drawn as dashed functions in figure 2.3 and the last column in table 2.1 lists the corresponding parameters. The GMM method found sufficient results for the data set.

Table 2.1.: Table of means μ and standard deviations σ for the three set Gaussians. The three functions' parameters are separated by dashed lines and the parameters are listed in the first column. The second column lists the parameters of the original Gaussians. The last column shows the results of the Gaussian Mixture Model (GMM) algorithm.

Parameter	A-Priori	GMM
μ_1	-1	-0.92
σ_1	1	1.1
μ_2	3	3.3
σ_2	2	1.9
μ_3	-6	-6.9
σ_4	7	7.1

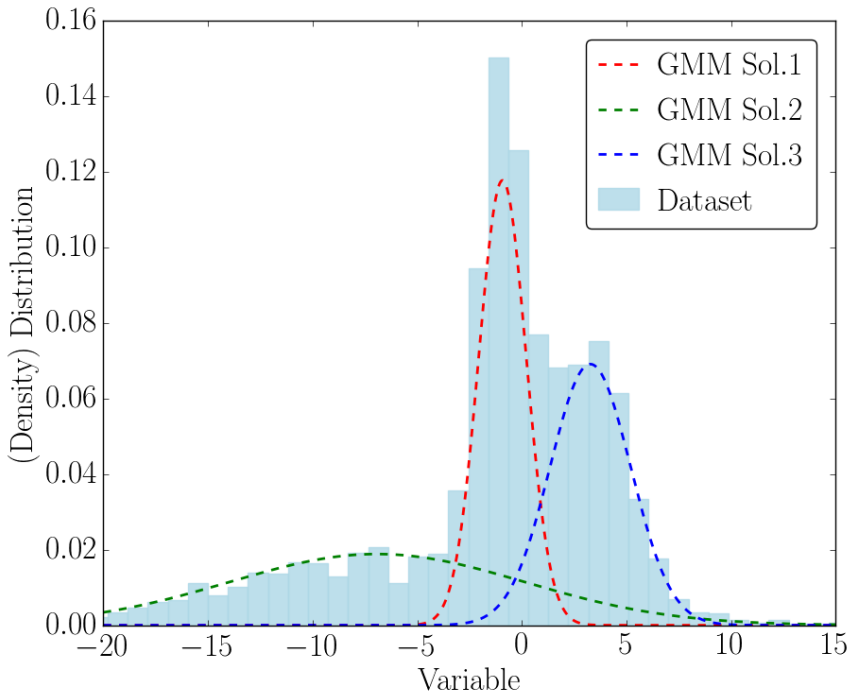


Figure 2.3.: Density distribution of an artificially generated data set. The blue histogram represents the generated data, based on three different Gaussians. The three dashed curves indicate the determined Gaussians by a Gaussian Mixture Model (GMM) algorithm. The parameters of all Gaussians (original and Gaussian Mixture Model (GMM) ones), can be seen in table 2.1

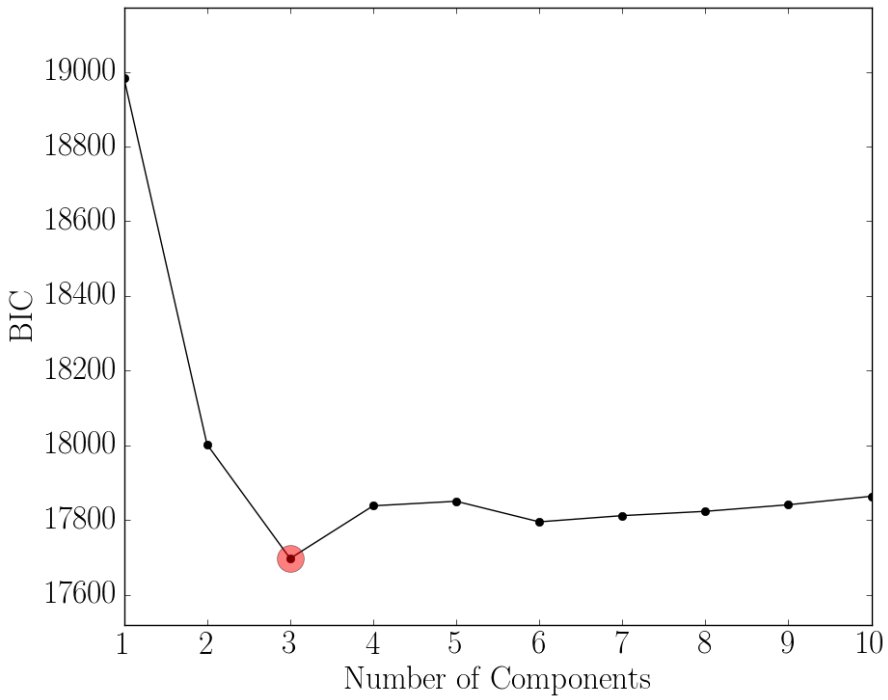


Figure 2.4.: Bayesian Information Criterion (BIC) curve of the Gaussian Mixture Model (GMM) algorithm that was applied on the data set shown in figure 2.3. The number of Gaussians was set from 1 to 10. The best Bayesian Information Criterion (BIC) value is indicated with a red circle and corresponds with a number of 3 Gaussian functions. The dots are connected with straight lines for better readability.

Clustering algorithms Besides box- and violin-plots, KDE's or mixture models, further unsupervised clustering algorithms are available. Miscellaneous methods have been developed and are used in this work to find the centre of clusters or to determine the number of possible clusters that cannot be described with Gaussian functions. These functions do not determine the shape or other statistical parameters but rather the centre or hull of (sub-) datasets.

The algorithm depends on the requirements: Is the number of clusters known or unknown? How many neighbouring data points are required to be associated with a cluster? Etc.

Most common algorithms are the Density-Based Spatial Clustering of Applications with Noise or K-Means++. The first one does not require a number of expected clusters. The number of neighbours that describe a cluster, as well as the minimum distance between two neighbouring points is necessary to run the algorithm. The second algorithm requires the number of clusters as an input and determines cluster centres that partition the data set in sub-sets in a way that the distortion (based on a certain metric) between the cluster centres and the associated data points becomes minimal. This procedure, called *Elbow Method*, is shown in the following two figures. Figure 2.5 shows 3 artificially generated data sets; the K-Means++ method computes the cluster centres for $k = 1$ to $k = 10$. For each result, the Elbow-Method is used to determine the distortion between the points and centres. Several implementations can be found in this regard: Normalised distance measurements between all data points within a cluster; distance measurements between data points and cluster centres, etc. Here, a normalised distance between the data and cluster centres is used. This is plotted on the y-axis of figure 2.6. The number of clusters is shown on the x-axis and for better readability, the values are connected with straight lines. Based on the Elbow-Method the ideal number of clusters is at the "elbow" of the curve, where the slope of the distortion flattens. Thus, a higher number of cluster points does not improve the clustering significantly, but results in an over-fitting. The elbow-point is indicated with a red circle and dashed lines indicate the approximate linear slope of the curve before and after the elbow.

The corresponding cluster centres are drawn as black circles in figure 2.5. For better visibility black dashed lines show the position. Red crosses show the actual mean values of the cluster that have been generated with three individual Gaussian functions. The actual and computed cluster centres fit perfectly. The colouring of the cluster data points is performed by the K-Means++ algorithm and shows that the separation was performed sufficiently.

Anyhow, the Elbow-Method has certain drawbacks. The shown data set is nicely separable with no background noise. Further, the overlap of the functions is not as severe as in figure 2.3. Recent studies suggest new metrics to determine the cluster performance. E.g., Tibshirani et al. [2001] introduced the *Gap-Statistics* that allows a more precise determination than the Elbow-Method.

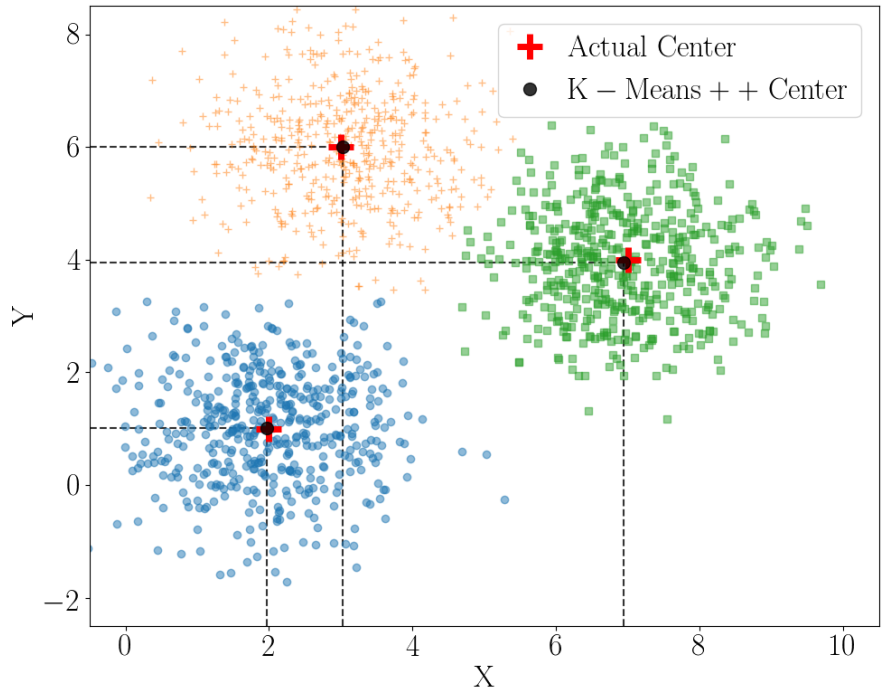


Figure 2.5.: Clustering of an artificially generated data set. Three clusters were generated, based on 2-D Gaussians, with certain mean values. These are indicated as red crosses. Black dots with dashed lines show the centres based on a K-Means++ algorithm and the data set points are shaped and coloured based on the clustering algorithm.

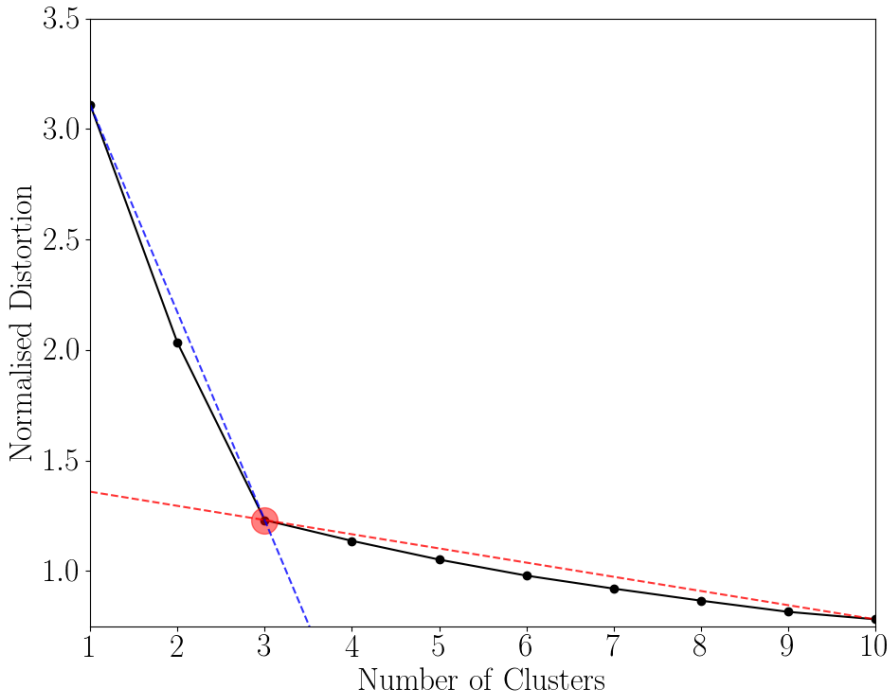


Figure 2.6.: Elbow method result of the K-Means++ algorithm for the data set shown in figure 2.5. The normalized distortion is plotted vs. the number of clusters and a red circle indicates the elbow point for 3 clusters. Dashed lines help as a guideline to show the slope change at this particular point.

2.1.2. Classification

This section introduces two methods for labelled (supervised) classification algorithms, the K-Nearest Neighbor (KNN) and SVM. Both methods are used in mass spectra classification projects on CDA data. Several methods are currently available with different degrees of complexity, computational performance and classification quality. However these methods are mostly used in the projects shown in chapter 5.

Furthermore, necessary pre-processing and feature extraction steps are not described. Note: several algorithms require certain input data properties, e.g., a mean of 0 and / or standard deviation of 1. Feature extraction cannot be described in general and depends on the use case: using the Fourier- or Wavelet-Transformation of a signal; using only the 5 highest peaks / frequencies etc. A few cases are shown later in the CDA chapter.

KNN Mathematically and from a programming perspective, the KNN algorithm is a rather simple supervised classification method. Consider figure 2.7, where two known data classes are present, namely P and Q (black squares and black circles, respectively). The grey lines indicate the hull of the data sets. An unlabelled datapoint DQ is classified based the distance to the nearest neighbours. In this example, the 3 nearest neighbours are used and the algorithm would assign DQ to group Q. Since a number of class corresponding neighbours can be determined, a probabilistic view on the classification can be performed, too (here: 2/3 assignment to Q, 1/3 to P). The hyper-parameter k that represents the number of neighbours, needs to be optimised with a training and validation set. With $k = 1$ each training point generates its so called Voronoi cell[2] and causes a severe over-fitting for a proper classification.

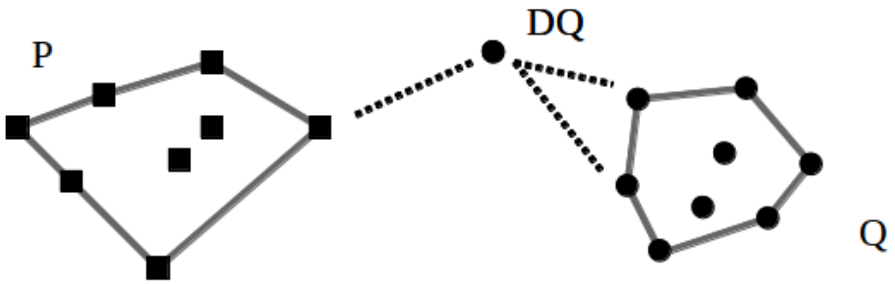


Figure 2.7.: Schematic visualization of a K-Nearest Neighbor (KNN) algorithm, where the point DQ is compared with the distance to the three nearest points w.r.t. the classes P and Q. The dashed lines show the shortest distances. From Albin [2014].

Although this algorithm appears simple, the crucial task for a data scientist is (besides the pre-processing and feature-extraction) to select suitable parameters. Besides k, other parameters need to be determined and tuned, too. The example shown in figure 2.7 uses apparently

[2]The Voronoi cell describes an area around a data point where the shortest distance, w.r.t. to all other data points, leads to the center of the cell. A resulting 2 dimensional Voronoi diagram reminds one of a tessellation / mosaic pattern.

an euclidean metric to determine the distances. However, depending on the task and the data structure (e.g., discrete or continuous) other metrics could perform better, like e.g. the Cityblock / Manhattan metric, the Post- / Paris-Trainnetwork-Metric, or other less common metrics like the Mahalanobis metric that is used e.g., for miscellaneous outlier detection algorithms.

SVM In den 1990s W. N. Wapnik and A. J. Tscherwonenkis developed in the *Bell-Lab*s the SVM. As part of the working group for adaptive systems, both set the algorithms and according learning theoretical concepts of this method [Cristianini and Shawe-Taylor, 2000, Chap. 4.2]. As the KNN method, most commonly used SVM implementations are used for supervised (multi-label) classification. There are also approaches for single class outlier identification and regression modelling.

The basic principle of (multi-dimensional) SVMs is the determination of a *Maximum Margin Hyperplane* and will be simplified summarised in the following. A complete and detailed derivation can be found in Cristianini and Shawe-Taylor [2000, Chap. 6] with applications in chapter 8 of the according reference.

First, we assume two data sets P and Q that contain two different classes, like e.g. a certain CDA mass spectrum and the remaining spectra types. Thus the complete data set is, where x_i is a n-dimensional data point assigned to one of the two classes:

$$X = P \cup Q = \{x_1, ...x_m\}$$

These two classes are shown exemplarily in figure 2.8, where black squares and dots indicate the classes P and Q, respectively. Gray connecting lines indicate the hull of the data sets.

Now, a hyperplane h needs to be determined that separates both data sets. A hyperplane is mathematically defined in the following, where w^{T} is the surface normal and b is the offset / bias of that plane. Note that w has the same (transposed) dimensionality as x.

$$h = \{x \in \mathbb{E}^n | w^T x = b\}$$

Since a hyperplane is needed to separate both data sets, the plane equation needs to be transformed in a way to distinguish between two classes. This is done in the following inequalities. A data point from P or Q shall result in a value larger or smaller than 0, respectively.

$$w^T x_i - b > 0,\ x_i \in P \tag{2.4}$$

$$w^T x_i - b < 0,\ x_i \in Q \tag{2.5}$$

Both inequalities can be merged in a single function that is always equal or larger than +1.

To fulfil this constraint, a variable y_i needs to be +1 or -1, depending on the corresponding class set of the data point x_i:

$$f(w, b, x_i) = y_i \left(w^T x_i - b \right) \geq +1 \tag{2.6}$$

So, depending on the class of a chosen x_i, the corresponding y_i needs to be chosen to fulfill this inequality.

Figure 2.8 shows two planes called $E1$ and $E2$ that both separate the training data accordingly. However, considering the point DQ that needs to be classified, a major problem can be identified. Although DQ is closer to the Q data set, $E2$ assigns the point to sub set P. $E1$ on the other hand appears more sufficient and separates both sets with a maximum distance d. This maximum width or maximum margin hyperplane is the theoretical concept of SVMs. To determine the corresponding parameters of this optimised plane the following method is applied.

The optimisation problem is done with so called Lagrange multipliers to find the ideal parameters for the hyperplane. The margin between two classes, based on the two classifiers from equation 2.4 and 2.5, can be determined and needs to be maximized. The term can be however transformed to $\|w\|^2/2$, which needs to be minimised.

In a Lagrange optimisation approach, the function that needs to be optimized is taken minus the side condition (here: equation 2.6). Since several data points x_i are available, the sum over all variables needs to be performed, times a Lagrange multiplier α_i, for each corresponding function $f(w, b, x_i)$. The resulting function that needs to be optimised becomes:

$$L(w, b, \alpha_i) = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i \left[f(w, b, x_i) - 1 \right] \tag{2.7}$$

Now, the equation 2.7 is derived to the according variables (partial differentiation). This leads to a set of linear equations that needs to be solved. The resulting values for the surface normal and offset are then the parameters that lead to the maximum margin hyperplane $E1$ as shown in figure 2.8. $E1$ is defined on three basis points, two from set P and one from set Q. These are the hyperplane corresponding support vectors and eponym for the machine learning method.

The original idea of the SVM is applicable for linearly separable data sets. However, separating the data linearly is not possible if the classes overlap or describe shapes that cannot be separated by a linear function. The first problem case can be handled with an additional "slack" parameter and a penalty parameter C. Data points, like outliers that are within a wrong class are "punished" and not considered in the according computations. The optimisation of this parameter is performed mostly via cross-validation.

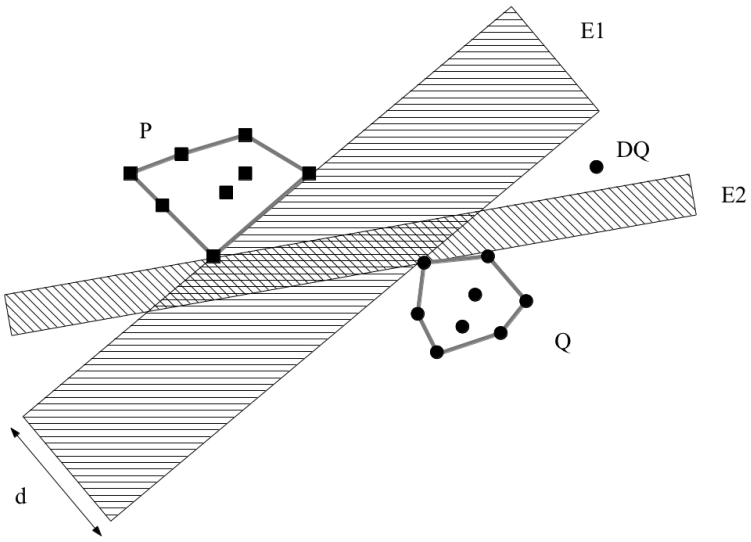


Figure 2.8.: Schematic visualization of a Support Vector Machine (SVM) algorithm. Two data classes are provided, namely P and Q. Both sets can be divided by an infinite number of hyperplanes. One example is $E2$ that assigns the unlabelled point DQ to set P, although its distance is shorter to the other set. With a Support Vector Machine (SVM) algorithm a hyperplane is found with the maximum width d that separates both classes ($E1$). DQ is consequently assigned to the set Q. Two points from P and one from Q are the support vectors of the plane. From Albin [2014].

The second case, shape depending inseparability, is handled with the so called Kernel-trick. With a so-called Kernel-function, the n-dimensional data set is put to a higher dimensional domain, where the data become linearly separable.

Both methods will be used in this work. A visualisation of the effect of the penalty parameter C and the kernel trick is shown in figure 2.9. Two data sets were generated with the class labels 1 and 2. Class 1 data are indicated with black triangles and class 2 data are shown as green circles. The data are given in 2-dimensional space. It can be seen that class 1 data are embedded within a surrounding class 2 set. In 2 dimensions, a linear separation is not possible, further, a few class 1 points are in the class 2 domain. A separation of both sets could be performed with a "circularly shaped" function. To achieve a separation a so-called radial basis function kernel, respectively RBF-Kernel is applied to transform the data to a higher dimension. The Kernel-function applies a radial Gaussian function on the data. Commonly used Kernels are polynomial or sigmoid functions. In higher dimensions a linear SVM is applied and then the data and decision function are transferred back to the original space for visualisation. The SVM decision function indicates the two class regions with a red and blue background colour. Due to a chosen penalty value a few class 2 data points are not considered. Further, the generated function appears to be suitable for further classifications of unknown data points.

Originally an SVM decision function defines strict boundaries. With data point weighting and additional modifications SVMs are also able to return a class and the corresponding probability.

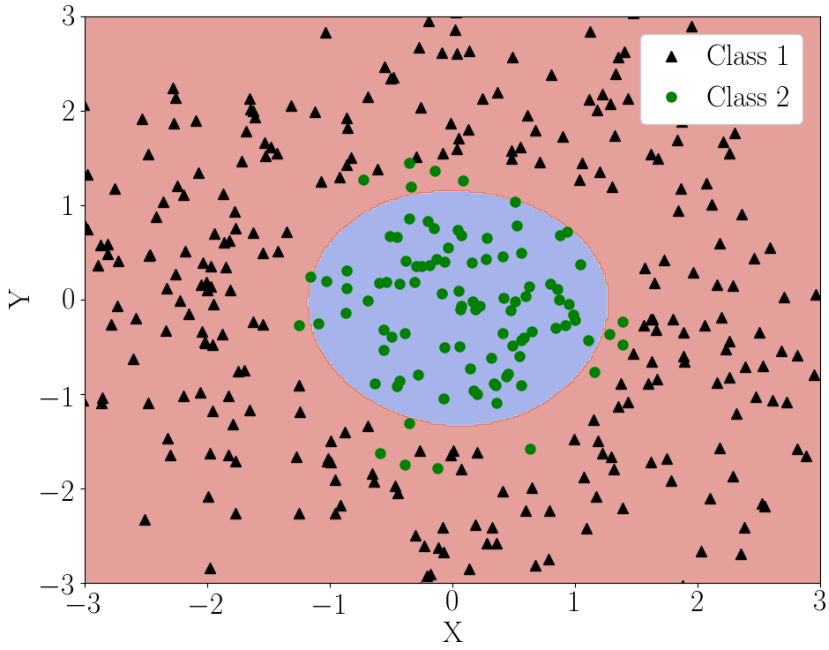


Figure 2.9.: Visualization of a kernel based Support Vector Machine (SVM) decision function. Two partly overlapping classes are available, indicated with black triangles and green circles. Since both classes are not linearly separable, a Gaussian kernel is applied to lift the data to a higher dimensional space, where the data become linearly separable. Then, a Support Vector Machine (SVM) is applied, with a penalty parameter C and then projected back on the original space. The resulting decision function is shown as a colour-map. Red and blue indicate the classification for class 1 and 2, respectively.

Performance

To validate the performance of classification algorithms, miscellaneous measures can be taken into account. Quantitative measures are for example the F1-score or the Odds Ratio.

A commonly used method is to visualise the classifications of the validation set in a so called confusion matrix. An example of an artificial classification result is shown in figure 2.10. The 2 times 2 matrix provides absolute numbers of the labelling. Vertically and horizontally are the results shown of the actual (a-priori) and predicted labels. Each matrix cell represents a number of classifications and a colour map is mostly used to improve the readability. This 4 cell matrix is the smallest possible matrix and used for binarised classifications (Class X - Not Class X). Multi-Class confusion matrices are accordingly larger. In total, the data set has 100 class 1 and 100 class 2 data points. 80 data points of the class 1 are predicted correctly, while 20 are labelled as class 2. The results for class 2 are worse. Only 65 of 100 are classified properly.

A confusion matrix appears simple, but several properties can be derived that determine the performance of the classificator. 80 of 100 particles are correctly classified Class 1 data. These *true positives* and *false negatives* can be used to determine e.g., the true positive rate and the false negative rate. Here, the according rates are 0.8 and 0.2, respectively, indicating a rather bad classifier. For the second row, the first entry shows the *false positives* and the second entry shows the *true negatives*. Here, according rates can be set for the negative class;

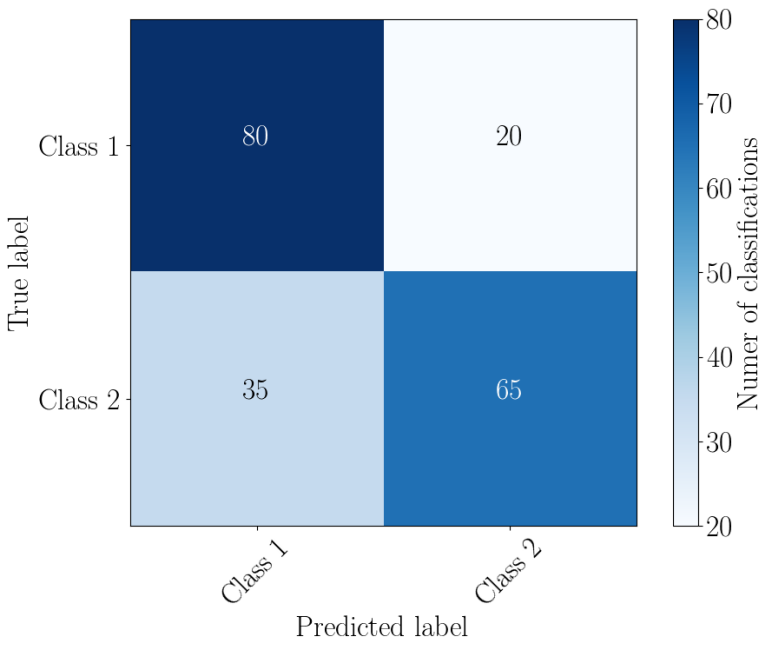


Figure 2.10.: Example of a confusion matrix for a binary classification problem. Numbers in each cell indicate the number of true positives, false negatives, etc. In total, 100 data points are class 1 and class 2, respectively. 20 of 100 class 1 data have been wrongly classified, as well as 35 of the class 2 set. A colour-map improves the readability of a confusion matrix, especially larger, multi-class ones. Normalized, a perfect classifier would result in an identity matrix.

class 2. Additional rates can be computed too, like e.g., the false discovery rate, or the negative predictive value. Further details can be found in the according literature and later in the CDA spectra classification section.

Based on the miscellaneous rates further performances can be derived and visualised. A common measure is the Receiver Operating Characteristics. This curve plots the true positive rate vs. the false positive rate for different parameters. The according integral varies then between 0 and 1. The worst case is a value of 0.5, indicating that for any parameter the number of correctly and wrongly determined measures stays the same. Consequently, the according classifier assigns the data with a chance of 50/50 to any class. With proper Receiver Operating Characteristics curves, an analyst can determine the ideal performance of a classifier, where the classifier performs well with a high true positive and according low false positive rate.

2.1.3. Neural-Networks

Besides classification and clustering problems, regression tasks are a third major topic in data analysis and machine learning. Commonly known and applied are linear regression models and algorithms that simply minimize the distance between linear fit and data points.

In this work one advanced method is used for data fitting that can also be applied for classification and other problems: Neural Networks. As mentioned in the introductory part of this chapter, the ideas are based on a first mathematical approach by Rosenblatt [1958], whose model used a simple single-perceptron to model a neuron. A graphical presentation of a single

neuron with n inputs is shown in figure 2.11. x_i input data are provided that are put with a weighting w_i, and an additional bias b to a so called activation function F. This function returns an output value y that can be used for further computations. The following list shows a fraction of possible activation functions. $F_{\text{perc}}(x)$ is the original idea based perceptron and returns only binary values. Further functions like the sigmoid (*Sig*) or the tangents-hyperbolicus (*Tanh*) functions return continuous values. Another commonly used function is the rectified linear unit that returns only positive values or 0 and is not sensitive for negative input parameters.

$$F_{\text{perc}}(x) = \begin{cases} 1 & \text{if } \underbrace{w \cdot x + b}_{z} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

$$F_{\text{sig}}(z) = \frac{1}{1 + \mathrm{e}^{-z}} \tag{2.9}$$

$$F_{\tanh}(z) = \frac{\mathrm{e}^{z} - \mathrm{e}^{-z}}{\mathrm{e}^{z} + \mathrm{e}^{-z}} \tag{2.10}$$

$$F_{\text{ReLU}}(x) = \max(0, x) \tag{2.11}$$

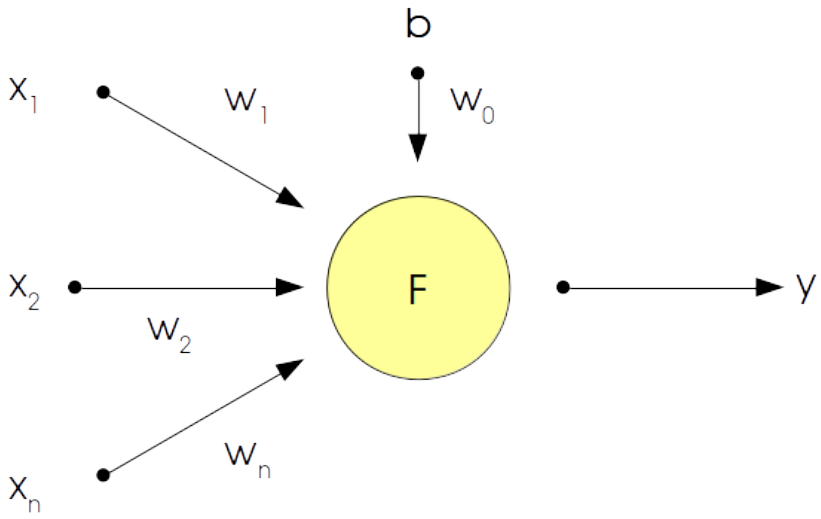


Figure 2.11.: Schematic overview of a single perceptron, respectively artificial neuron. The activation function F requires input data x_i, multiplied with a certain weight w_i. Additionally, a bias / offset b can be set, too. Depending on the activation function, a value y is returned, representing e.g., a class label, discrete value, probability, etc.

Choosing an activation function depends on the computational task and data structures. *Tanh* and *Sig* functions are commonly used for classification problems. $F_{\text{sig}}(z)$ returns con-

tinuous values between 0 and 1 and can be sufficiently used for probability computations and output, while $F_{\tanh}(z)$ is also sensitive to negative values and returns a corresponding negative output. A binary problem with two classes uses these functions to distinguish mathematically between two differently labelled data. rectified linear units provide a continuous output, depending on x and further modified versions of this function return also negative data (like e.g., so called leaky rectified linear units).

The theoretical background of a single neuron can now be used to generate Neural Networks as exemplary shown in figure 2.12. The structure of such a network can be divided in 3 parts:

Input Layer
: The input layer is the first layer of the network and obtains the input data. It is directly connected outside its network structure.

Hidden Layer
: The hidden layer is not interacting with any user input. Hidden layers are neurons that are connected to each other. So the first hidden layers get output values from the input layer as an input. The generated output of the hidden layer are then send to an output layer or to another layer of hidden neurons.

Output Layer
: Finally, the last layer returns the output values for the data scientific problem. E.g., a regression value, or a set of classification labels with corresponding probabilities. At the very end, the user needs to evaluate and interpret the data for further analysis.

The shown network simplifies the structure of actual neural network applications. The number of layers, especially for deep-learning networks, can be vast and also a variation of different activation functions can be applied. Further, different architectures have been identified and developed that fulfil certain tasks more accurately than others. For example so-called convolutional networks appear to work sufficiently for image data. These networks search for patterns and structures in images, independent of the patterns' location within the image matrix. Other concepts like Long Short-Term Memory Cells have a certain "memory" of time-depending data and are used currently e.g. for speak-recognition. These recurrent neural networks have been developed, where hidden-layers reconnect with previous hidden layers. A detailed description of all possibilities, advantages, disadvantages etc. would be too extensive for this work[3]. Also adaptive networks can be developed that will also be used in this work. A so called *Dropout-Rate* determines how often neurons are activated, and what contribution these neurons provide. Less activated neurons, are then deleted during the training process to optimize the performance and computation time. These Dropout-Rates can also be set randomly, too.

[3] A recent compilation of currently used networks with according literature and applications can be found here: `http://www.asimovinstitute.org/neural-network-zoo/`, 05.03.2018. A detailed list and schematic overview of the basic architectural concepts are shown.

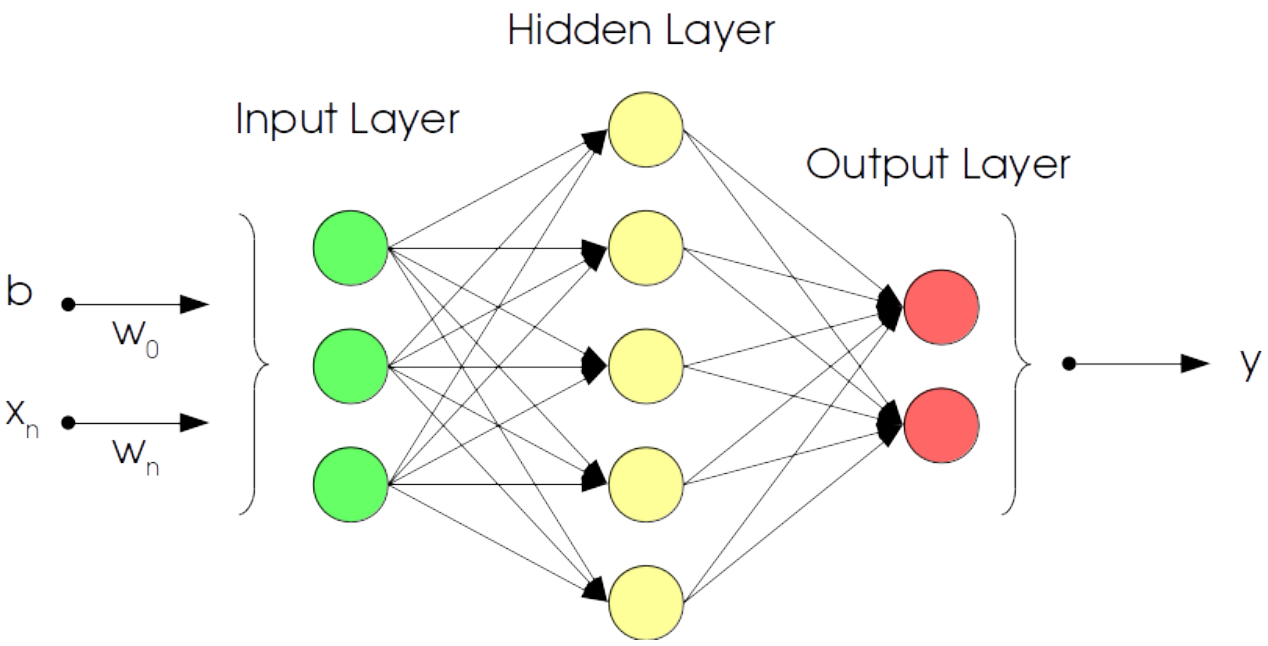


Figure 2.12.: Schematic overview of an artificial Neural Network (NN). The Neural Network (NN) is divided in three layers: input and output layer are the "communication layers", where a user can input data or obtain results. Hidden layers are layers within these two communication layers and can have complex structures and depths. A variation of activation functions, network architectures etc. is possible and depends on the computational task. Based on the bias b and the input x_n a value y is returned, e.g. a classification label or regression value. Several output values are also possible, like e.g. multi-label classes with corresponding probabilities.

The training of a neural network optimizes the different weighting w_i and biases / offsets b for each function, for each neuron of each layer (input, hidden and output). Considering a single perceptron as shown in figure 2.11 the optimization is rather straightforward. First, initial small weightings are chosen and the result y for a certain input set is computed. The resulting error between output and expected label or regression value is then used to adjust w accordingly to the difference between prediction and actual value. Since neural networks combine several layers with miscellaneous features and activation functions, more sophisticated algorithms need to be used.

Several methods have been developed and a common way of optimization is the so called *Back-Propagation*. Further, this method is related to the gradient descent method of finding an optimal solution: Based on the derivative of a loss function, for each iterative computational step, parameters are changed to find the minimum of the loss function.

The Back-Propagation uses first initial weighting and offset parameters and a set of input data is send to the neural network. At the end of the network, so after the last output layer, the outputs are compared with the actual expectation. A deviation between computed / predicted value and actual value is defined as the error of the neural network. Then, weights are adjusted from the output to the input layers. Based on a scalar, the learning rate, and the derived error / loss functions w.r.t. to the according weights, the weights are adjusted slightly. Then, input data is send again through the NN. This "sending data forward" - "adjusting weights

backwards" method is done until the error tolerance goes below a user set value. To prevent oscillations around a local minimum solution, the learning rate can be adjusted adaptively.

A sophisticated and detailed description as well as several NN applications with TensorFlow can be found in Bonnin [2016].

2.2. Monte-Carlo simulations

The determination of physical laws has been historically done with rather simple experiments, where observed data would be fitted with theoretical expectations. Or, these data lead directly to the corresponding equation describing the nature of the experiment, like e.g., the discovered Photo-Effect by Einstein that lead to the energy equation of a photon namely $E = h \cdot f$ with h as the Planck constant and f as the frequency of the photon.

With more computational power and the ability to calculate numerical simulations, the link between input data and output law or output statistics became more complex. Non-symmetric distributions, non-linear equations or a chain of extensive equations lead to computational problems that cannot be computed analytically. In these cases, Monte-Carlo (MC) computations are performed. The idea behind this method is to use randomly chosen parameters and values (based on assumptions, a-priori knowledge and input distributions), to determine the statistical output or expectation of the computational problem.

The categories of problems and the algorithms and methods for a MC simulation are vast and depend on the requirements based on the problem. In this work, MC algorithms are described in detail in the according sections. Since MC simulations are used in this thesis to determine the orbital elements of cosmic dust particles, the used methods can be generally described as in the following:

1. An instrument records observational data of a cosmic particle (video image / in-situ dust detector)

2. The mean / average observational parameters and the corresponding error(s) are determined

3. Based on these values for each observed particle a large set of possible observational solutions is computed

4. All generated particles are then used to determine miscellaneous astro-dynamical properties

5. At the end a large set of different solutions is computed for each dynamical property. These distributions are then used to determine final statistical solutions (e.g., with KDEs or box plots)

In the following a classical example is shown that visualizes a simple MC problem and the computational method. Assume a 2-dimensional domain, where $x \in [-1, 1]$ and $y \in [-1, 1]$. Within the domain is a circle defined with a radius of $r = 1$. Thus, the ratio between an inscribed circle and the corresponding square is $\pi r^2/4r^2$; the ratio becomes $\pi/4$. Now, (e.g., within a for-loop) randomly generated points are computed within the domain, thus the maximum value per dimension varies between -1 and 1. Now, the number n of generated points is kept per computational iteration. Further, per iteration a counter m is raised, if a point is within the circle, so the euclidean distance to the centre is equal or less than 1. The ratio of m/n is expected to correlate with the circle and square area. Thus, π can be computed with this MC method via: $\pi = 4 \cdot (m/n)$. With more iteration steps, the MC simulation shall converge with π.

The result of this MC method is shown in figure 2.13. The y-axis indicates the computed value, where a horizontal dashed line indicates the value for π. 11 MC simulations have been performed with 1 Million iteration steps (x-axis). The median of the solutions is shown as a black solid curve. It can be seen that the curve converges with π. At $n = 10^6$ the computed value for π is $3.1409\ldots$, compared with the first digit values of $3.1415\ldots$.

This underscores that (depending on the problem though) MC computations require a large number of (re-)samples to perform sufficiently.

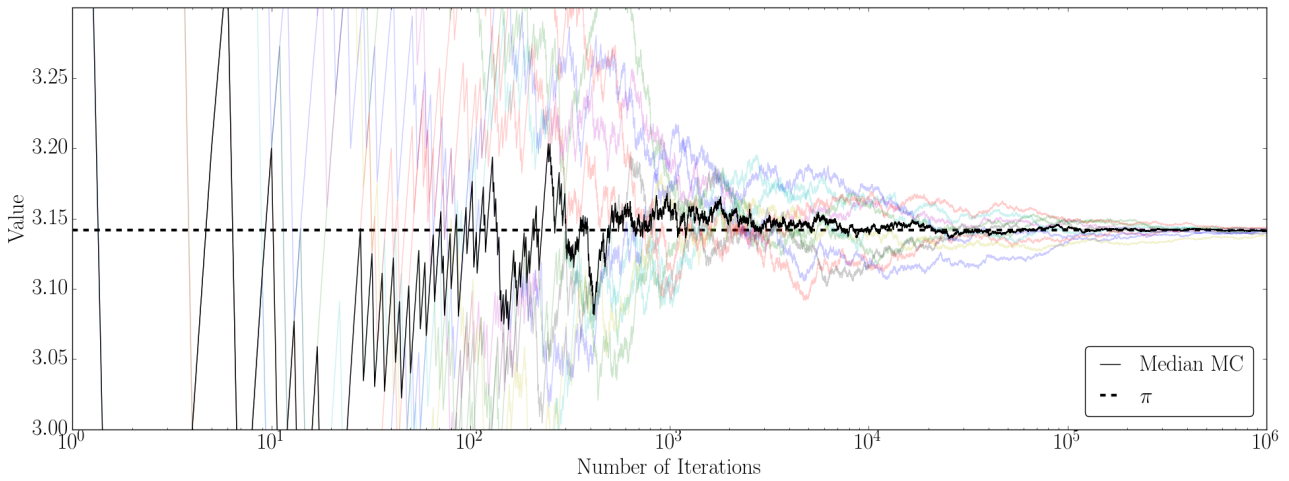


Figure 2.13.: Visualization of a Monte-Carlo (MC) computation to determine the value for π. The determined value is plotted vs. the number of computational iterations (method is further described in the text). 11 times 1 Million iterations are performed (coloured curves) and the black curve shows the median of these results. A black dashed line indicates the value for π. It can be seen that for large iteration steps, the Monte-Carlo (MC) results converge with π.

3. Orbit Dynamics

This theory chapter describes and defines the concepts that are used in the astro-dynamical part of this work. For the Canary Island Long-Baseline Observatory (CILBO) and the Cosmic-Dust-Analyzer (CDA) databases have been developed that contain astro-dynamical information of meteor and in-situ detected dust particles. Since the astro-dynamical field is vast, proper parameters need to be chosen for future scientific and technical analysis.

First, the basic concepts of astro-dynamics are shown that describe the orbital elements and reference frames. Additionally, the Sphere Of Influence (SOI) and Hill Sphere are explained. These conceptional spheres are used to simplify the orbit computation codes that are described in the following chapters. Then, radiant maps are explained. The description is linked to the reference frame section, thus it is recommended to work through the entire chapter. Furthermore the Tisserand parameter and D-Criterion are introduced in the section regarding dynamical linkage. In a final short part the in-situ measurement depending pointing effects are described, like e.g., the Kepler-Ram.

Most dynamical computations, the determination of directional information or the usage of different reference frames are based on the National Aeronautics and Space Administration Spacecraft Planet Instrument C-matrix Events (SPICE) toolkit[1]. A sophisticated and detailed documentation can be found in the according sub-directories of the website. Further, since this work focuses on the implementations and usage of PYTHON, the SPICE wrapper SPICEYPY was used [AndrewAnnex]. A detailed description of SPICE would exceed the scope of this work. However, kernels and other SPICE related files that have been developed in order to perform computations are shown in the appendix.

Theoretical concepts in astrodynamics are extensive. For a sophisticated and detailed overview on dynamics Murray and Dermott [2000] shall be considered. Jenniskens [2008] applies and introduces several methods for meteor research. Furthermore, this work considers only mechanical solutions. Radiation effects etc. are not considered to properly determine "as is" orbits.

3.1. Orbital Elements

Two objects that revolve around a united centre of mass describe, in a restricted two body problem with no further disturbances, two orbits. These orbits can be described analytically by

[1] `https://naif.jpl.nasa.gov/naif/index.html`, 12.02.2018

using the Keplerian / Orbital elements and can be derived using classical mechanical laws. Historically a first qualitative description was provided by Johannes Kepler, the *laws of planetary motion*:

1. Planets revolve around the Sun in an ellipse shaped orbit
2. The Sun is at the centre of one focal point
3. The area speed is constant

The later derived orbital elements describe the shape, orientation and alignment of the orbit and are commonly used in planetary science. Furthermore, the orbital elements (using no external disturbing forces) are time independent. The movement of a planetary object is given by 6 elements, 3 position elements (x, y, z) and the corresponding velocities $(\dot{x}, \dot{y}, \dot{z}) = (v_x, v_y, v_z)$. Additionally, the time T is necessary too, to determine the actual position of the object itself. Thus, to describe only the shape of the orbit 5 elements are required that are explained in the following. Figure 3.1 and 3.2 illustrate the described elements. Some elements are defined with respect to a reference plane. In the Solar System the ecliptic plane in J2000 is commonly used (see section 3.2). The ecliptic plane is spanned by the X-Y plane of the ecliptic J2000 reference frame, where the X-Axis is pointing towards the vernal equinox Υ (sometimes, but uncommon: γ). The Z-Axis is pointing towards the ecliptic north, thus perpendicular to the plane that is spanned by the Earth's motion around the Sun. The Y-Axis is derived from the right-hand-rule and points consequently to the summer solstice. J2000 indicates that the corresponding time frame is the year 2000. The orbital elements are:

a Semi-major axis, given in meters. The semi-major axis describes the longest radius of the ellipsoidal shaped orbit. The pendant is the semi-minor axis, normally abbreviated with a b.

e Eccentricity. The eccentricity describes the deviation of an orbit from a perfectly circularly shaped orbit. The following definitions apply:

$e = 0$	Circle
$0 < e < 1$	Ellipse
$e = 1$	Parabola
$e > 0$	Hyperbola

i Inclination, given in degrees. The enclosed angle between the orbital plane and the reference frame. The following definitions apply:

$0° \leq i < 90°$	Prograde Orbit

$i = 90°$ Polar Orbit

$90° < i \leq 180°$ Retrograde Orbit

$\Omega, ☊$ Longitude of ascending node, given in degrees. This angle defines the angular distance between the node and the vernal equinox. At the node's position, a planetary body intersects with the reference frame (moving from the south to the north).

ω Argument of periapsis, given in degrees. This parameter sets the angular distance between the periapsis point and the ascending node of the orbit.

Based on the semi-major axis and eccentricity the periapsis q and apoapsis Q can be computed (Note: no apoapsis can be computed for $e \geq 1$): $q = (1 - e)a$ and $Q = (1 + e)a$. To determine the position of an object on the orbit two additional parameters are needed:

M Mean anomaly, given in degrees. The mean anomaly describes the angular distance between the periapsis and the body, moving on a conceptual circular orbit with a constant speed. (The angular distance of the actual position on the true orbit w.r.t. the periapsis is the true anomaly ν that is shown in figure 3.1).

T_{J2000} Epoch, given in seconds passed J2000 (commonly used). This value is the Mean anomaly corresponding time.

Besides the commonly used orbital elements, further elements are sometimes used like the longitude of descending node ☋ that is 180° shifted from the ascending node. The longitude of perihelion Π, given in degrees, is mostly used in meteor science, especially to compute the D-Criterion between particles (see section 3.5). This parameter is simply defined as $\Pi = \omega + \Omega$.

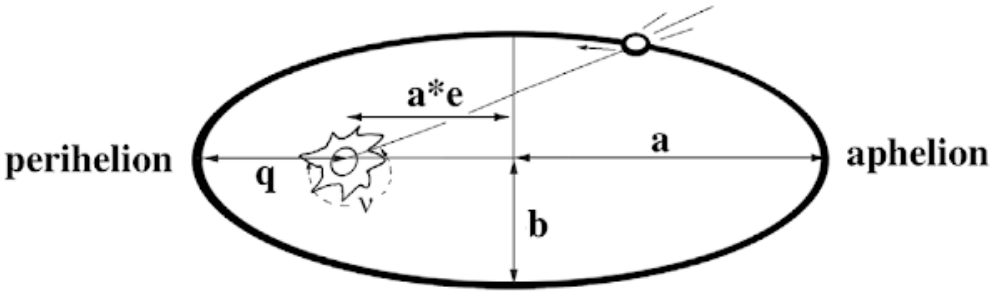


Figure 3.1.: Top view on the orbital plane of a comet (top, on orbit) revolving the Sun (left, illustrated with sun rays). The Sun is located in one focal point of the ellipse with the semi-minor and semi-major axes b and a, respectively. The eccentricity is e and q is the perihelion distance. The aphelion is indicated on the right. ν shows the true anomaly of the comet and is drawn closely below to the Sun. From Jenniskens [2008].

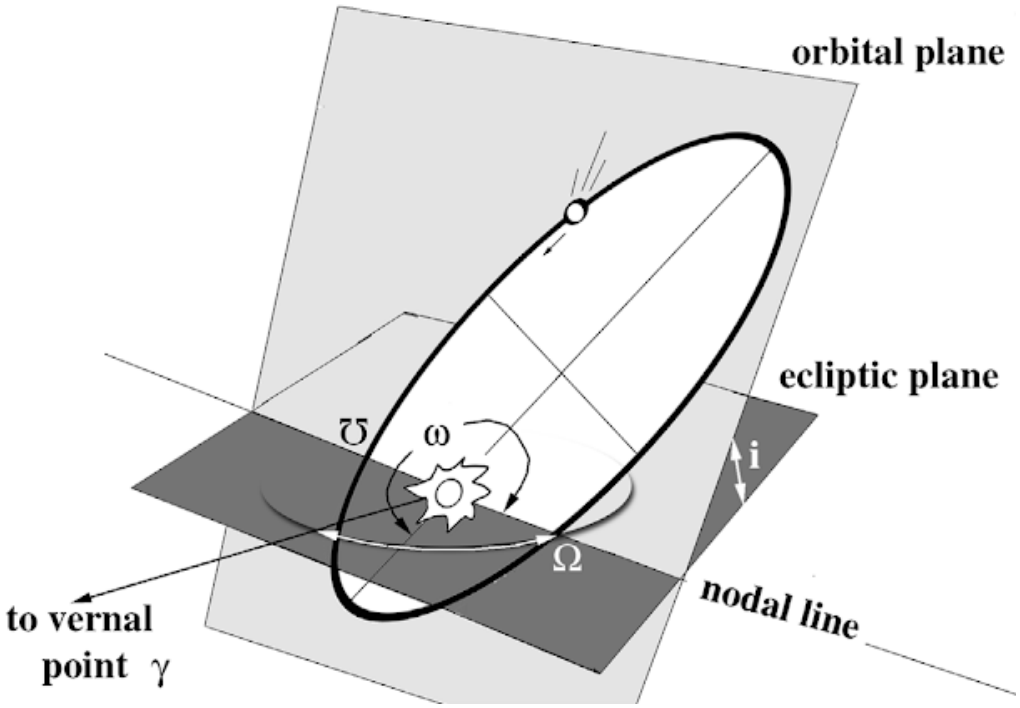


Figure 3.2.: 3-dimensional view of a comet's orbit. The orbital plane from figure 3.1 is aligned with respect to the ecliptic plane and encloses the inclination i. Furthermore, the longitude of ascending Ω and descending $\mho$ nodes are shown along the nodal line. The vernal equinox' symbol is here a γ. ω shows the argument of periapsis. From Jenniskens [2008].

3.1.1. Computation

In meteor research and in-situ dust observation, one obtains 3 observational parameters that can be used as a proper input for dynamical computations. First, the actual position vector $\vec{r}$, given in an arbitrary reference frame. Second, the corresponding velocity vector $\dot{\vec{r}} = \vec{v}$; and the corresponding time information T_{J2000}, given in seconds passed J2000:

$$\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{3.1}$$

$$\vec{v} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \tag{3.2}$$

The norm of both vectors will be given as $r = |\vec{r}|$ and $v = |\vec{v}|$.

The gravitational constant $G = 6.672 \cdot 10^{-11}\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}$ is given, and the mass of the major M and minor m objects are needed. These parameters are needed when applying Newton's law of physics. A more convenient abbreviation of the masses and G is the parameter $\mu = G(M+m)$. Assuming $m \ll M$: $\mu \simeq GM$.

The semi-major axis a can be derived directly from the Vis-Visa Equation [Murray and Dermott, 2000, p. 30-31]:

$$\boxed{a = -\frac{\mu}{2} \cdot \left(\frac{v^2}{2} - \frac{\mu}{r} \right)} \tag{3.3}$$

For the other elements Bate et al. [1971, chap. 2] provides a detailed and sophisticated overview of the derivation. Here, a summary is given.

Further computations require the determination of the angular momentum vector $\vec{h}$. Its norm is simply $h = \left|\vec{h}\right|$.

$$\vec{h} = \vec{r} \times \vec{v} \tag{3.4}$$

$\vec{h}$ is aligned consequently perpendicular to the orbital plane. Now, an auxiliary vector needs to be defined that is parallel to the nodal line as seen in figure 3.2. This can be achieved by a cross product of a vector that is perpendicular to the reference frame and $\vec{h}$. The vector $\vec{n}$ is introduced and defined as (norm: $n = |\vec{n}|$):

$$\vec{n} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \times \vec{h} \tag{3.5}$$

The eccentricity of an orbit is derived from the norm of the eccentricity vector $\vec{e}$. This vector points from the center of mass towards the periapsis of an orbit. The magnitude of the vector is the corresponding eccentricity:

$$\boxed{e = \left| \frac{1}{\mu} \left(\left(v^2 - \frac{\mu}{r} \right) \vec{r} - \langle \vec{r}, \vec{v} \rangle \cdot \vec{v} \right) \right|} \tag{3.6}$$

The inclination angle i is the angle enclosed by the orbital plane and the reference frame. Consequently, i is the same angle that is enclosed by the angular momentum vector $\vec{h}$ and its Z-component vector $\vec{h_z} = (0, 0, h_z)^{\mathrm{T}}$ that is perpendicular to the reference frame. Now, the dot product can be used that is defined as $\vec{h} \cdot \vec{h_z} = \left|\vec{h}\right| \left|\vec{h_z}\right| \cos i$. The left side of the equation leads to $\vec{h} \cdot \vec{h_z} = h_z^2$. The equation becomes $h_z^2 = h \cdot h_z \cos i$. Consequently, the inclination can be computed:

$$\boxed{i = \arccos \frac{h_z}{h}} \tag{3.7}$$

The same procedure can be used to determine the longitude of ascending node. However, here, the enclosed angle Ω is between the vectors $\vec{n}$ and $\vec{n_x}$, where $\vec{n_x} = (n_x, 0, 0)^T$ and is aligned with the reference point direction (e.g., the vernal equinox):

$$\boxed{\Omega = \arccos \frac{n_x}{n}} \tag{3.8}$$

The argument of periapsis is defined as the angle between the periapsis direction and the longitude of ascending node. Consequently, the dot product between the eccentricity vector $\vec{e}$ and the nodal vector $\vec{n}$ leads to ω:

$$\boxed{\omega = \arccos \frac{\vec{n} \cdot \vec{e}}{ne}} \tag{3.9}$$

Determining the angle between $\vec{e}$ and the actual position vector of an object $\vec{r}$ leads to the true anomaly $\nu = \arccos \langle \vec{r}, \vec{e} \rangle / re$.

One method to determine the mean anomaly M is:

$$M = E - e \cdot \sin E \tag{3.10}$$

E is the eccentric anomaly and describes the angle between the semi-major axis of an orbit, with a vector on a projected circular orbit that has a radius equal the semi-major axis. The vector points towards a conceptual point on the circle that is projected from the actual orbit (projection: vector from semi-major axis aligned perpendicular to the semi-major axis, pointing to the object's position. See figure 2.7 in Murray and Dermott [2000]). Since the equation is not analytically solvable, only a numerical approach can be applied. The eccentricity anomaly can be determined by using the following equation (further detail see Murray and Dermott [2000, p. 32ff]):

$$E = \arccos \left(\frac{1}{e} - \frac{r}{ea} \right) \tag{3.11}$$

3.2. Reference Frames

This thesis analyses partly the particle dynamics from Earth-bound and Saturn in-situ dust measurements. The orbital elements (and further properties like the radiants, see section 3.4) are given with respect to a centered body and the angular measures, like the inclination refers to a defined reference frame. For this work, three reference systems are defined that are explained in this section.

Equatorial J2000 This is the most commonly used reference frame for Earth-bound observations. The X-axis of the frame is the vernal equinox direction of the Sun, as seen from Earth and the Z-axis is the North Pole direction of the rotation axis. Thus, Earth's equator draws the reference plane dividing the latitudes in positive and negative values, respectively. The latitude is defined as the *declination* δ in degrees and the longitude is the right ascension α, where the angles are given in hours (from 0 to 24 h).

The equatorial J2000 system will be mostly used for the meteor research part, where the radiants (see section 3.4) of meteor showers are provided in this coordinate system. Furthermore, this system is one of the default system that is provided by SPICE.

Helio-centric ecliptic coordinates (HCEC) The HCEC system is similarly defined as the equatorial system. The X-axis points towards the vernal equinox, but the Z-axis is directed to the ecliptic North Pole of the Solar System w.r.t. J2000. The north-south pole axis of the ecliptic is perpendicular to the ecliptic plane that is drawn and defined by Earth's path around the Sun.

This system provides angular positions in longitude l and latitude b, both given in degrees. This Sun- or Solar System barycentre-centered frame is commonly used to determine the state vectors of planets or interplanetary space missions.

Saturn-centric non-rotating coordinate system (SCNRC) For the Cassini mission and to analyse CDA measurements in a geometric context the SCNRC system is defined. This system is Saturn-centric and the Z-axis points towards the rotational North Pole of Saturn. The coordinates of this axis (in J2000) are:

$$\alpha = +40.58^\circ$$
$$\delta = +83.54^\circ$$

Commonly in SPICE, a coordinate system is defined by two vectors. The primary axis (here: Z-axis) is fixed set. The secondary axis, here the X-axis, points "as close as possible" to the vernal equinox. Since the coordinate system's primary axis (Z-axis) corresponds with the axial tilt of Saturn w.r.t. to the ecliptic plane, the X-axis of the coordinate system cannot point

directly to the vernal equinox. The axis aims the vernal equinox' corresponding meridian line in ecliptic coordinates. The geometric configuration can be seen in figure 3.3.

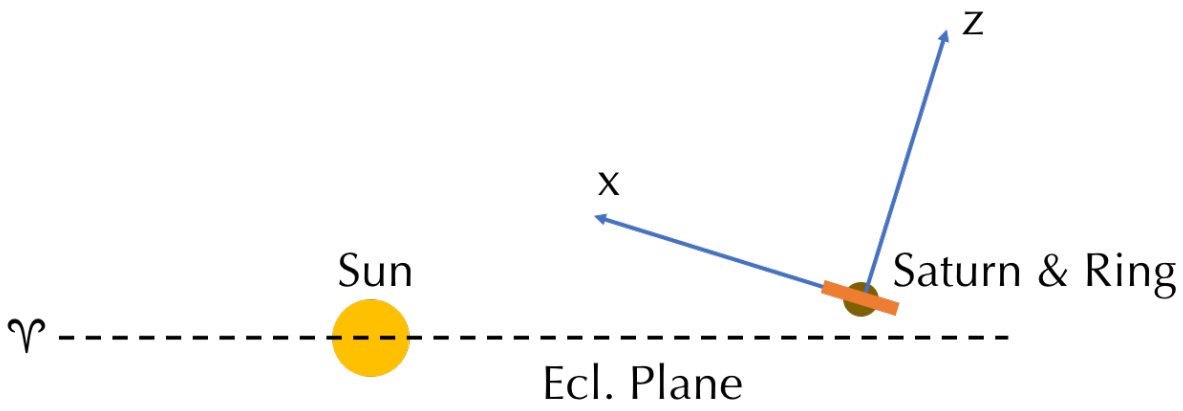


Figure 3.3.: Definition of the Saturn-centric non-rotating coordinate system (SCNRC). The schematic plot is a side view of the Sun (yellow), Saturn (brown) at Earth's autumn equinox and the ecliptic plane (black dashed line). Due to its inclination, Saturn is not drawn on the plane. The Z- and X-axis of the Saturn-centric non-rotating coordinate system (SCNRC) are shown. Since the vectors need to be perpendicular, the X-axis cannot point directly to the vernal equinox that is allocated left from the Sun.

This coordinate system does not belong to the standard SPICE repository and needs to be defined manually. The kernel information can be found in the appendix A.1

Non-Inertial reference frames As will be later shown, miscellaneous meteor streams can only be identified using a rotational coordinate system that is aligned to the Sun [Jenniskens, 2008, p.499, chapter 28, 29 and 31]. This coordinate system is used for Earth-centric observations, and will be called Earth-centric rotating coordinate system (ECRC). A similar Saturn-centric system that co-rotates with its orbit will be called Saturn-centric rotating coordinate system (SCRC). Both corresponding SPICE kernels are not in the standard repository, thus need to be defined manually and are shown in the appendix A.1 (listing A.2 and A.3, for Saturn and Earth, respectively).

Figure 3.4 gives a schematic overview of the frame definition. The figure shows the Solar System from a top-view, with the Sun in the centre and three different positions of a planet (brown). The corresponding orbit is drawn as a black dashed half-circle. The X-axis is pointing from the planet to the Sun and the Z-axis points towards the ecliptic North Pole (not shown). Both axes are considered as primary and secondary axis in a rotating reference frame. The Y-axis direction results form the right-hand-rule and points towards the Anti-Apex direction of the planet. The movement of the planet itself is indicated on top with a red vector (v). Different longitude values are named in table 3.1.

Table 3.1.: Labels / Names of directions in a rotating reference frame.

Longitude in degrees	Label
0	Helion-Direction
90	Anti-Apex Direction
180	Anti-Helion Direction
270	Apex Direction

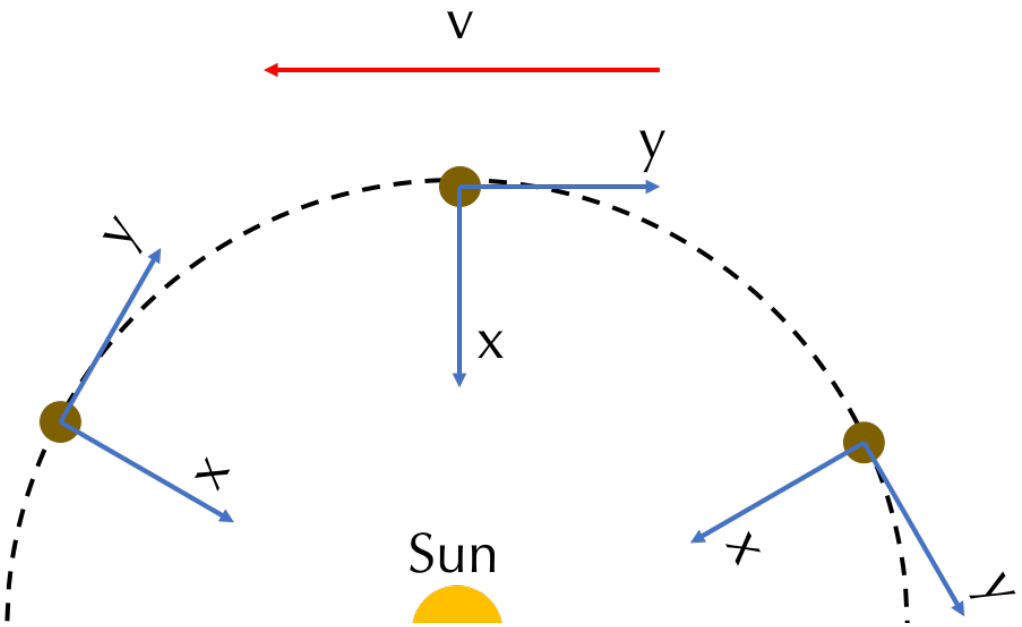


Figure 3.4.: Definition of the Saturn-centric rotating coordinate system (SCRC) and Earth-centric rotating coordinate system (ECRC). The schematic plot shows the Solar System from a top-view. The Sun is shown in the centre and the dashed line indicates the orbit of a planet (brown). In this use case it is Earth or Saturn. The movement of the planet is indicated on top with a red vector (v). Three positions of the planet are shown with the corresponding, rotating reference frame. The X-axis is pointing towards the Sun, the Z-axis points towards the ecliptic North Pole (not drawn), and the Y-axis results from the right-hand-rule and points to the Anti-Apex direction of the planet.

3.3. Gravitational Spheres

To simplify dynamical simulations, three- or multi-body problems are reduced to two body problems by considering which nearby body is the gravitational dominant factor. These simplified solutions allow then faster analytical computations rather than extensive numerical integrations. This can only be applied if one neglects secondary radiation effects, or long-term effects like resonances.

Two concepts are available for a restricted two body problem, the Hill-Sphere and SOI. Within the Hill-Sphere, a satellite body that is embedded in a larger gravitational system (e.g. Earth around the Sun) and that is significantly smaller and lighter than the major body can revolve in a stable orbit. The Lagrange points L1 and L2 are allocated at the border of this

sphere. The radius r_{Hill} of the Hill-Sphere is defined in the following. Here, m is the mass of the major planet (e.g., Earth) and M is the mass of the larger body (e.g., Sun). Both, Hill-Sphere and SOI are depending on the distance between the two bodies. However, since most objects that are considered in this thesis are on an almost perfectly circular orbit, the semi-major axis a is used:

$$r_{\text{Hill}} = a \cdot \left(\frac{m}{3M}\right)^{1/3} \tag{3.12}$$

A detailed derivation and further theoretical background can be read in Murray and Dermott [2000, section 3.13].

A second gravitational sphere concept is the Sphere Of Influence (SOI). Although the detailed description computes an oblated spheroid, this work uses the simplified spherical solution. Within the SOI the movement of a smaller body can be computed in a reference frame w.r.t. to the centre body. This is commonly used in space engineering to determine analytically the energy / momentum transfer on a spacecraft for a fly-by manoeuvre. For this work, the SOI is used to simplify interplanetary dust dynamics by reducing the trajectory computations to a two-body problem. Bate et al. [1971, page: 333] describe the usage of this concept for the Apollo missions and defines the radius of the SOI as:

$$r_{\text{SOI}} = a \cdot \left(\frac{m}{M}\right)^{2/5} \tag{3.13}$$

A second definition can be found in Messerschmid and Fasoulas [2017, section 4.6.2], who define the radius as:

$$\tilde{r}_{\text{SOI}} = a \cdot \left(\frac{m}{M}\right)^{1/2} \tag{3.14}$$

For this work the SOI definition by Bate et al. [1971] is used. And further details on miscellaneous radii can be read e.g. in Albin [2014].

3.4. Directional Information

Based on the velocity vector of a meteor or in-situ measured particle, the so called radiant can be computed. The radiant gives the source, respectively origin coordinates of a particle in any reference system. This parameter is determined by projecting the upstream direction of a particle on a unit sphere. For Earth-bound meteor observations this is done in equatorial J2000 coordinates. Meteor showers like the Geminids in December or the Perseids in August describe a cluster of radiants at a particular sky region. Furthermore, to sort out non-stream meteors

with the same radiant information, the in-falling velocity norm is considered, too. Table 7 in Jenniskens [2008] provides a detailed list of known day- and night-time meteor showers with their corresponding average properties. For example: The Perseids peak in mid August at a radiant position α, δ with the velocity V_{obs}:

$$\begin{aligned} \alpha &= 3 \text{ h } 10 \text{ m} \\ \delta &= 58^\circ \\ V_{\text{obs}} &= 59 \text{ km/s} \end{aligned}$$

Given these parameters for a set of meteors further (minor) streams can be identified, e.g., by using 3-D wavelet transformation algorithms as described in recent works by Brown et al. [2010]. Further stream detections methods are shown in section 3.5. Recent numerical approaches to describe the streams have been performed by Soja et al. [2015c,b].

Besides these shower meteors, random, or sporadic meteors collide also with the Earth. Using a co-rotating reference frame (see section 3.2) like the ECRC, leads to a map of sporadic streams (see Jenniskens [2008, chap. 28]). Figure 3.5 provides a radiant map of the sporadic meteors coming from Earth's Apex direction, the Helion and Anti-Helion direction (streams from high elliptic comets / asteroids) and the Anti-Apex. Two more sources are present, namely the Toroidal sources that are associated with highly inclined streams. Probably, the shown significance is due to an observational bias of radar observations; see Jenniskens [2008, Chap. 29].

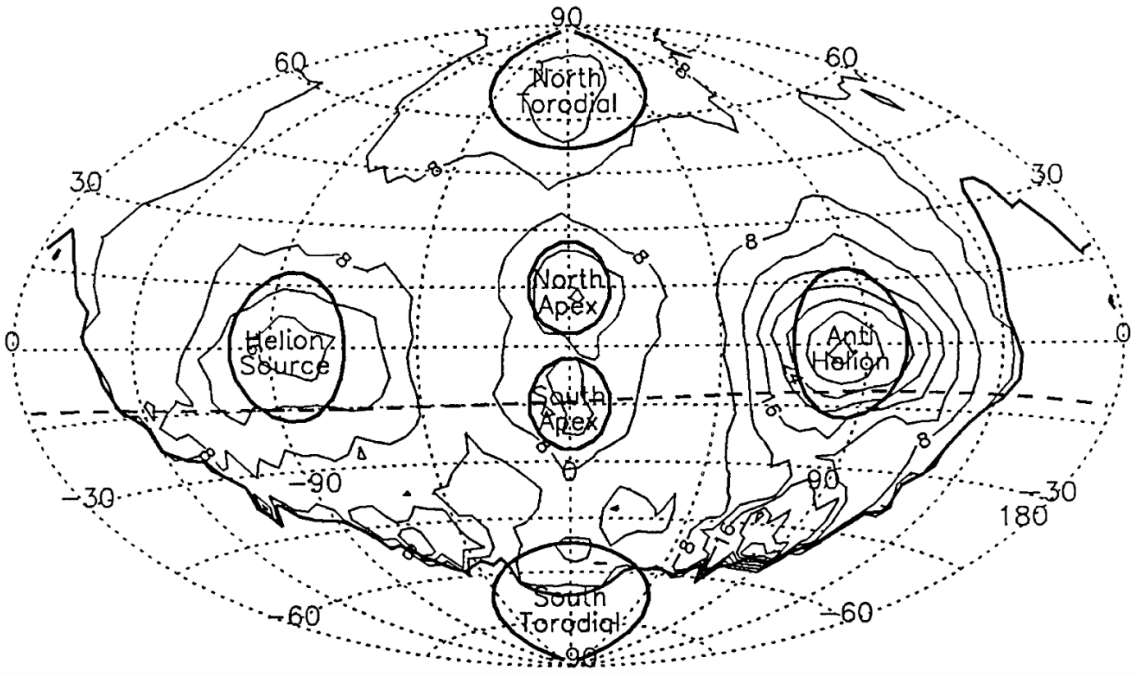


Figure 3.5.: Radiant map of sporadic meteor sources. Only meteors with a mass of at least 10^{-4} g are considered. Further details on the data and the figure are in Taylor and Elford [1998].

3.4.1. Zenith Attraction

Giovanni Schiaparelli found already in the late 19$^{\text{th}}$ century a bias effect on meteors called the *Zenith Attraction*. Due to the gravitational pull of the Earth, the trajectories are bend towards Earth's centre of mass. Consequently the velocity vector at the atmosphere entrance is, compared with the velocity vector at the SOI entrance, more shifted towards the zenith. A short summary and graphical derivation can be found in Andreev [1991].

In geocentric coordinates, the azimuth angle is not affected. Only the elevation angle of the radiant is biased. The difference between the radiant's elevation at Earth and at Earth's SOI is computed in the following equation, where ΔZ indicates the elevation shift. V_{obs} is the observed velocity, and V_{inf} is the velocity at infinity; Z_0 is the apparent angle between the radiant and the zenith.

$$\Delta Z = 2 \cdot \arctan\left(\frac{V_{\text{obs}} - V_{\text{inf}}}{V_{\text{obs}} + V_{\text{inf}}} \cdot \tan\left(\frac{Z_0}{2}\right)\right) \tag{3.15}$$

Based on the observed velocity V_{obs}, V_{inf} can be easily computed by the following equation, where G is the gravitational constant, M and R are the mass and radius of the major object and h is the height above the surface of the major object, where the particle was detected:

$$V_{\text{inf}} = \sqrt{V_{\text{obs}}^2 - \frac{2GM}{R+h}} \tag{3.16}$$

Meteors on Earth occur at an average altitude of around 100 km above sea level. Thus, for meteor science, the equation can be simplified to:

$$V_{\text{inf}} \simeq \sqrt{V_{\text{obs}}^2 - 123\ \frac{\text{m}^2}{\text{s}^2}} \tag{3.17}$$

For in-situ measurements in the Saturnian system, the zenith attraction applies also for the Interplanetary Dust Population (IDP) measurements and particle detections from Interstellar Direction (ISD). Here, a general value for h cannot be provided, since in-situ measurements are possible at any distance from Saturn.

A second method to un-bias the zenith attraction are shown in the sections regarding the orbit code for CILBO and CDA, respectively. Simplifying the orbit dynamics into a restricted two body problem can be performed, to analytically determine the intersection point of a particle with the major body's corresponding SOI (for Earth- and Saturn-based observations). Computing the radiant with the velocity vector at the SOI's border considers and un-biases the gravitational pull of the major object. The equations shown above are supportive if no proper trajectory can be determined.

However, for an analytical quick analysis, the zenith attraction can be computed by only using the values V_{obs}, h and Z_0, replacing V_{inf} in equation 3.15 with equation 3.16. For meteors on Earth the simplified equation 3.17 can be used instead of 3.16.

3.5. Dynamical Linkage

Computing orbital elements for a large set of cosmic dust particles leads to a vast database with multidimensional dynamical properties. Based on the radiant and in-fall velocity, one can roughly determine major streams and possible dust source regions, like the ISD particles. A proper analysis of the inner-dynamical linkage of particles, the linkage of particles with major bodies, or the identification of possible parent bodies and (major / minor) streams require a more sophisticated analytical approach.

This section describes two quantities that are used to determine relations and linkages between cosmic particles and larger bodies. First, the Tisserand Parameter is introduced that is derived from Tisserands criterion and can be used to determine the dynamical relations between a minor and major body in a three body problem. Second, the D-Criterion is introduced, especially the modified [Jenniskens, 2008, p. 596 + 597] D_{SH} parameter [Southworth and Hawkins, 1963]. This parameter is commonly used in meteor research to determine streams. This parameter might also be useful for further in-situ analysis.

3.5.1. Tisserand Parameter

To distinguish between asteroids and comets, miscellaneous empirical and analytical approaches were performed. E.g., in meteor science, to identify "asteroid-like" or "cometary-like" orbits, Whipple [1954] introduced the K-Criterion that takes the semi-major axis and eccentricity into account. Kresák [1967] defined the Pe-Criterion as a product of the period time and eccentricity. Further, he introduced also the Q-Criterion that is simply the determined aphelion. The most common approach is the usage of the Tisserand Parameter T, defined by François F. Tisserand.

The Tisserand Parameter is derived from Tisserand's relation. Based on the restricted three body problem and the derived constant Jacobi-Integral, Tisserand found that the semi-major axis, eccentricity and inclination are conserved in a way. Example: A comet has the orbital elements a, e and i. After a close encounter at Jupiter the comet gains or losses energy / momentum. Consequently, the new parameters are $\tilde{a}$, $\tilde{e}$ and $\tilde{i}$. In this three body problem (third body is the Sun), taking into account that $m_{\text{Sun}} + m_{\text{Comet}} \approx m_{\text{Sun}} + m_{\text{Jupiter}}$, the following relation applies:

$$\frac{1}{2a} + \sqrt{a(1-e^2)}\cos i = \frac{1}{2\tilde{a}} + \sqrt{\tilde{a}(1-\tilde{e}^2)}\cos \tilde{i} \tag{3.18}$$

With the Tisserand relation planetary bodies can be identified as the objects, whose orbital elements were altered due to a close fly-by with a major body like Jupiter.

Based on the relation above, the Tisserand parameter is defined as:

$$T_{\mathrm{P}} = \frac{2a_{\mathrm{P}}}{a} + 2\cos i\sqrt{\frac{a}{a_{\mathrm{P}}}(1-e^2)} \tag{3.19}$$

The index P is a place holder for the corresponding major object, or planet. To apply this function, the orbit of the major body needs to be approximately circular. Further, the gravitational influence on the minor body is mostly affected by the major body and the Sun.

With respect to Jupiter, the Tisserand parameters are separated in three groups:

Asteroids Asteroids in the main belt between Mars and Jupiter have a Tisserand Parameter $T_{\mathrm{J}} \geq 3$.

Jupiter family comets Jupiter family comets are comets with an aphelion at around 5 Astronomical Unit (AU), respectively at Jupiter's orbit. After a close encounter of a Halley' Comet with Jupiter, its orbital elements are changed (Tisserand relation applies). The comet revolves around the Sun afterwards in a closer orbit and is dynamically linked with Jupiter. A Tisserand parameter between $2 < T_{\mathrm{J}} < 3$ indicates such an object.

Damoclids Named after the object 5335 Damocles. These objects have $T_{\mathrm{J}} \leq 2$, and revolve the Sun typically in a high eccentric orbit, with a large semi-major axis (aphelion beyond Jupiter).

With the discovery of active asteroids, a finer differentiation was introduced and named that also considers the presence of a coma. Further details can be found in Jewitt et al. [2015, Fig. 1].

Enceladus One of the major rings in the Saturnian system, the E-Ring, is feed by particles from Enceladus (details on in-situ measurements in the Saturnian system can be found in chapter 5). The emitted particles have been dynamically studied in miscellaneous theoretical and numerical approaches [Horanyi et al., 1992, Beckmann, 2008]. Since the eccentricity of Enceladus is around 0.005 and the mass of Saturn is significantly larger than the mass of the moon, the Tisserand parameter can be used for dynamical analysis studies. Since other moons revolve around Saturn that are even larger than Enceladus, the computed Tisserand parameter in SCNRC needs to be considered carefully. The Tisserand parameter applies best for a three body problem with less perturbations, like a comet that revolves around the Sun and is perturbated by Jupiter. Further Tisserand parameters w.r.t. to other moons can be determined too, like, e.g., Titan. Since Enceladus supplies the Saturnian system with particles and interacts with the E-Ring one can expect promising dynamical analysis with this particular

moon. For example, the Tisserand parameter could reveal a strong dynamical linkage between the dust population and Enceladus.

Beckmann [2008] determined numerically a sophisticated model and distribution of the dynamics of the E-Ring. He computed the density of miscellaneous elements in phase diagrams, like e.g., the eccentricity vs. the semi-major axis [Beckmann, 2008, Fig. 5.9, p. 73]. To identify possible links to the Tisserand parameter, figure 3.6 is additionally computed for this work. The color map indicates the Tisserand Parameter (filtered between 2 and 3), depending on the eccentricity and semi-major axis, given in Saturn Radius / Radii (SR). For simplification, the inclination is set to 0 degrees. The intransparent black area excludes solutions that fall into Saturn and the gray transparent area shows the solution space that intersects with the A-Ring. Inclined particles however could survive between the gaps of the rings.

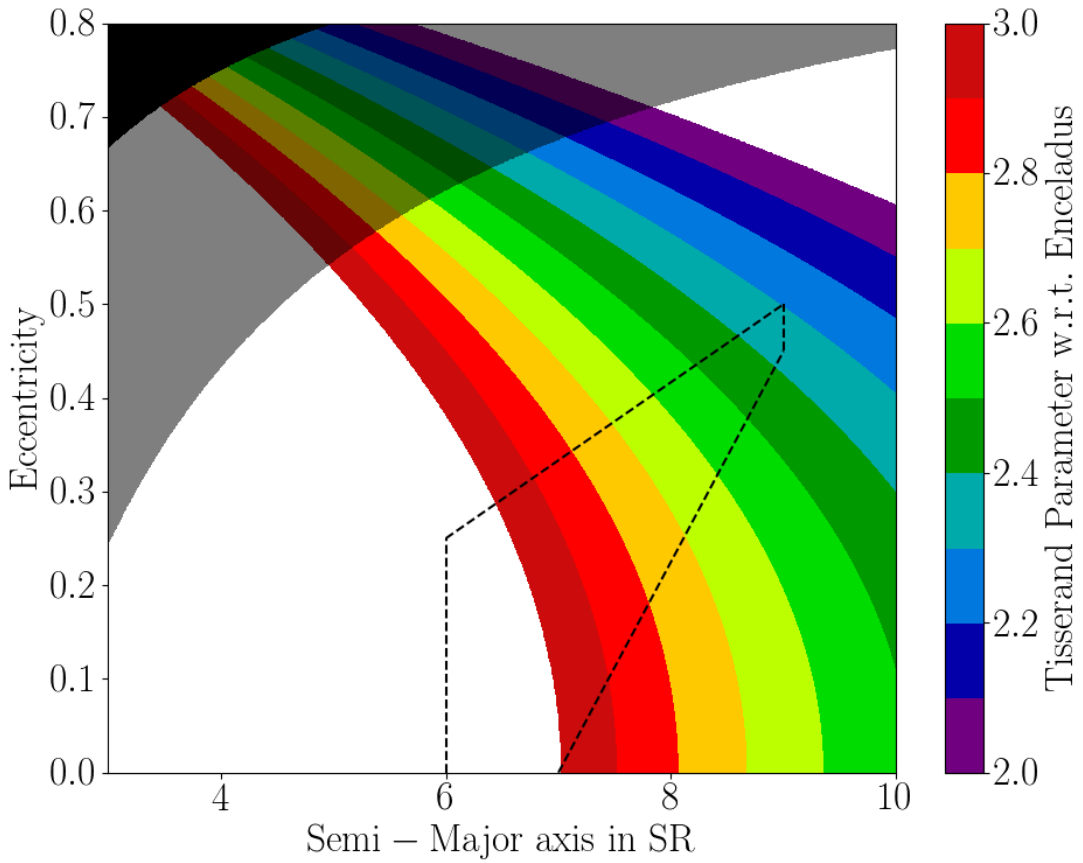


Figure 3.6.: Tisserand solution space for Saturn-bound particles with respect to Enceladus. The plot shows a color map, indicating the Tisserand parameter for particles with $i = 0°$ in Saturn-centric non-rotating coordinate system (SCNRC). The Y- and X-axis show the eccentricity and semi-major axis, given in Saturn Radius / Radii (SR). The black, intransparent area shows the solution space, where the periapsis is smaller than 1 Saturn Radius / Radii (SR). The grey, transparent area indicates the solution space, where the periapsis is within the A-Ring. A small area is enclosed by dashed black lines and gives approximately the solution space of numerical computations [Beckmann, 2008, Fig. 5.9, p. 73] for E-Ring particles with a radius less than 500 μm.

Dashed black lines enclose a solution space, where Beckmann [2008] determined a dense distribution for E-Ring particles with a size smaller than 500 μm. A major part of the simulated

particles is within a Tisserand Parameter range of 2 to 3, indicating a dynamical linkage to Enceladus. Larger particles however have mostly a Tisserand Parameter larger than 3; revolving Saturn in a large circular orbit [Beckmann, 2008, Fig. 5.9 (top right), p. 73].

The Tisserand parameter is successfully used to distinguish between different asteroid and comets types in the Solar System. Further, this parameter might be useful for other dynamical use cases, like in-situ measurements within a planetary system. The Saturnian system consists of several larger moons that supply and disturb the planetary dust, namely Titan, Enceladus and also Phoebe.

3.5.2. D-Criterion

The Tisserand Parameter is a sophisticated and analytically derived way to determine the astro-dynamical linkage between a small body, and a major one, both embedded in a restricted three body problem. The Tisserand Parameter however, cannot be used to link particles together. Based on the radiant and in-falling velocity meteor streams can be determined on Earth. A dynamical approach that considers the orbital elements was introduced by Southworth and Hawkins [1963]. They defined the so called D_{SH} parameter to distinguish between streams, to identify new (minor) streams and / or to associate an individual meteor with a group of other measurements.

The parameter takes the following elements into account: eccentricity e, periapsis q, inclination i, the argument of periapsis ω, and the ascending node and node of periapsis, Ω respectively Π:

$$D^2 = (e_1 - e_2)^2 + (q_1 - q_2)^2 + \left(2\sin\left(\frac{i_1 - i_2}{2}\right)\right)^2 \tag{3.20}$$

$$\begin{aligned} D_{\mathrm{SH}}^2 =& D^2 \\ &+ \sin(i_1)\sin(i_2)\left(2\sin\left(\frac{\Omega_1 - \Omega_2}{2}\right)\right)^2 \\ &+ \left((e_1 + e_2)\sin\left(\frac{\Pi_1 - \Pi_2}{2}\right)\right)^2 \end{aligned} \tag{3.21}$$

Jenniskens [2008, p.479 - 482, 596] provides an overview on how to handle and interpret the D_{SH} parameter properly. To identify streams or possible groups of similar particles, the D_{SH} parameter is computed for each possible particle pair. n particles lead then to a symmetric $n \times n$ matrix with the according D_{SH} values. The elements on the diagonal are 0, since the orbital elements are the same. It is recommended to compute the matrix and compare the possible mean or median values (depending on the underlying distribution) of the stream with possible parent bodies.

Particles that are grouped in a major stream or meteor shower have a mean D_{SH} value of

0.06 or less while minor streams have a value up to 0.17. Particles that are associated in a group shall have a D_{SH} value of less than 0.20.

The D_{SH} parameter works well for low inclined particles of less than 30 degrees. For higher inclinations the results are less significant. Modifications of the D_{SH} parameter have been done by miscellaneous researchers. Besides the parameter by Southworth and Hawkins [1963] the following parameters are introduced:

D_{D} By Drummond [1981]

D_{H} Jopek [1993]

D_{N} Valsecchi et al. [1999]

The performance of all available D-Criteria was analysed and compared e.g. by Galligan [2001], who found that all criteria classify at least 70 % of stream meteors in the according group.

In this work, the commonly used D_{SH} parameter is used. Chapter 5 refers to two works by Bötsch [2018] and Faber [2018], who apply the D_{SH} on in-situ measurements by Cassini's CDA.

3.6. Pointing geometries

Optical instruments in meteor research and in-situ dust instruments in space have entirely different technical requirements. Most meteor stations are installed steadily, pointing towards a sky area that is fixed in azimuthal coordinates with respect to the location of the camera. Depending on the location, pointing direction and the camera set-up miscellaneous bias effects alter the flux measurements and orbit determination of the in-falling dust populations.

For in-situ measurements, pointing geometry effects can be determined easily. In HCEC or SCNRC the velocity of the spacecraft and dust particles are of the same order of magnitude. Thus, the pointing of the dust instrument needs to consider the relative speed between the spacecraft and particle. Figure 3.7 shows this schematically. A spacecraft SC moves in a reference frame with a certain, frame centred velocity v_{SC}. Further, a particle moves in the system with a velocity v. To optimize the according boresight of an instrument on-board the spacecraft (bs), the particle's velocity vector needs to be considered in spacecraft-centric coordinates (inverted red dashed line). The resulting angle α between the boresight and the spacecraft velocity vector results in an optimal impact angle between the instrument and incoming particle (anti-parallel impact direction w.r.t. the boresight vector in spacecraft-centric coordinates). The expected flux direction as seen from the spacecraft (so velocity corrected) is called Dust-Ram and needs to be known for an optimised in-situ observation campaign.

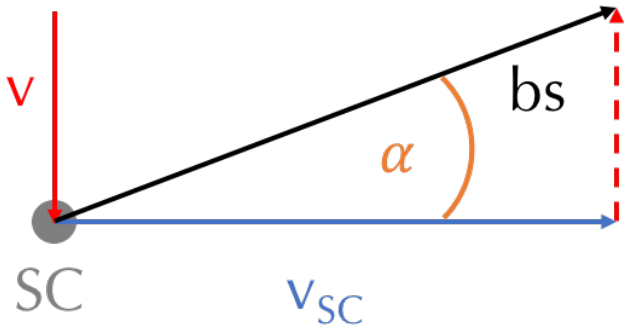


Figure 3.7.: Schematic overview of the boresight pointing correction / optimisation. A particle (red arrow, velocity v) and a spacecraft (grey, SC with velocity v_{SC}) propagate in a certain reference frame. To observe the incoming particle or particle flux from this particular direction, the boresight of an instrument on-board the spacecraft needs to be set, depending on the velocity. The optimum is an anti-parallel impact velocity vector w.r.t. the boresight in spacecraft-centric coordinates (bs).

Kepler-Ram Since the dynamical properties of the particles are unknown, one needs to set assumptions regarding the possible incoming direction, depending on the possible dust source directions. In section 3.4 the radiants of miscellaneous dust sources have been described, e.g., the Apex or Anti-Helion dust source as well as dust of interstellar origin. Dust models like the E-Ring by Beckmann [2008] can be used as an input to determine possible directional fluxes in the Saturnian system.

A simplification of bound particle orbits are orbits with $e = 0$; i.e. perfectly circularly shaped. The velocity v of these particles is constant and is described by the formula below, where M is the mass of the central body, G is the gravitational parameter with $G = 6.67259 \cdot 10^{-11}$ $\mathrm{m^3 kg^{-1} s^2}$ and d is the distance between the particle and the major body. Since the particle's mass is several orders of magnitude smaller than the major body's one ($m << M$), the particle's mass is not taken into account:

$$v \approx \sqrt{\frac{GM}{d}} \tag{3.22}$$

The Dust-Ram of these $e = 0$ particles is called Kepler-Ram and can used in planetary systems like in the Saturnian system, or for IDPs in HCEC.

Plasma-Ram The Saturnian system has an extended magnetosphere, interacting with Saturn's moons and the helio-centric magnetosphere. Figure 3.8 shows a detailed drawing of Saturn's magnetosphere, its properties and interactions. The magnetic field lines co-rotate with the rotation of Saturn with an angular frequency of $\omega = 2\pi/38,362.5$ Hz (10.656 hours rotation period). Charged particles couple with the field lines, co-rotate with the field, induce currents etc. The composition of Saturn's plasma revealed miscellaneous components, like protons and hydronium [Smith et al., 2008]. Beyond 9 SR protons are the dominant component

[Young et al., 2005]. The inner part of the Saturnian system is especially generated by particle emissions of Enceladus.

These plasma clouds that are captured in the field can enter in-situ dust instruments like the CDA. They induce charges and signals in the grids and targets and generate an impact event that is not related with a dust impact. Such false positives are classified as noise or, wrongly, as an actual impact. For possible bias analysis the Plasma-Ram needs to be determined. The orbit velocity v is approximately the co-rotational velocity (note: for larger distances a factor of 0.7 to 0.8 times the co-rotational speed should be applied for a rough estimate):

$$v = \omega \cdot d \tag{3.23}$$

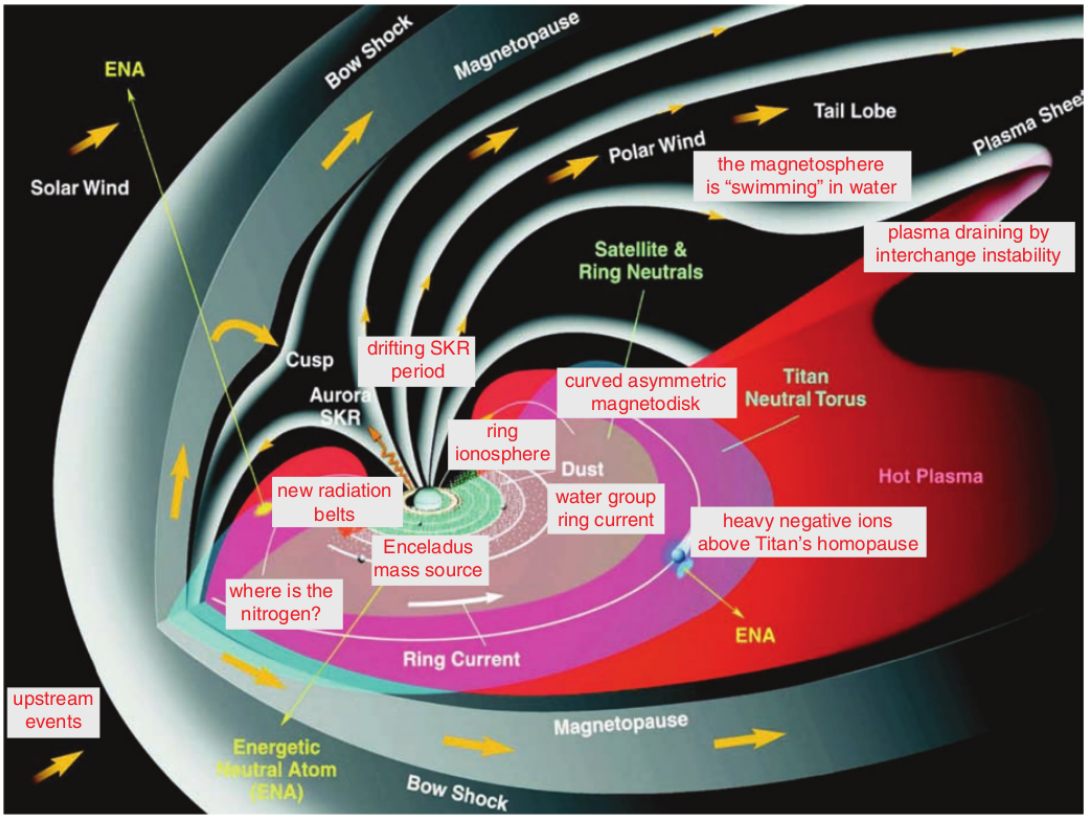


Figure 3.8.: Schematic drawing of Saturn's magnetosphere. From Dougherty et al. [2009].

However, an extensive analysis and summary of Saturn's magnetosphere by Dougherty et al. [2009] shows that magnetic field coupled charged particles are affected by neutral gas clouds that are also present in the system. This causes a drag force; consequently, the particles do no co-rotate perfectly with the magnetic field. All Plasma-Ram computations assume a reduced velocity by 30 %. Figure 3.9 shows the Kepler-Ram, co-rotating, and Plasma-velocity in km/s vs. the radial distance in SR. The co-rotational and Plasma-Ram velocity intersect with the Kepler-Ram velocity between approximately 1.86 and 2.36 SR. In a noise bias analysis this should be taken into account, since Kepler- and Plasma-Ram measurements could not be

distinguished. The angular distance between Kepler- and Plasma-Ram cannot be taken into account: Due to large the angular Field-of-view (FOV) of CDA (Impact Ionisation Detector (IID): around 50 degrees opening angle), a possible Plasma-induced error is always presents when observing the Kepler-Ram direction.

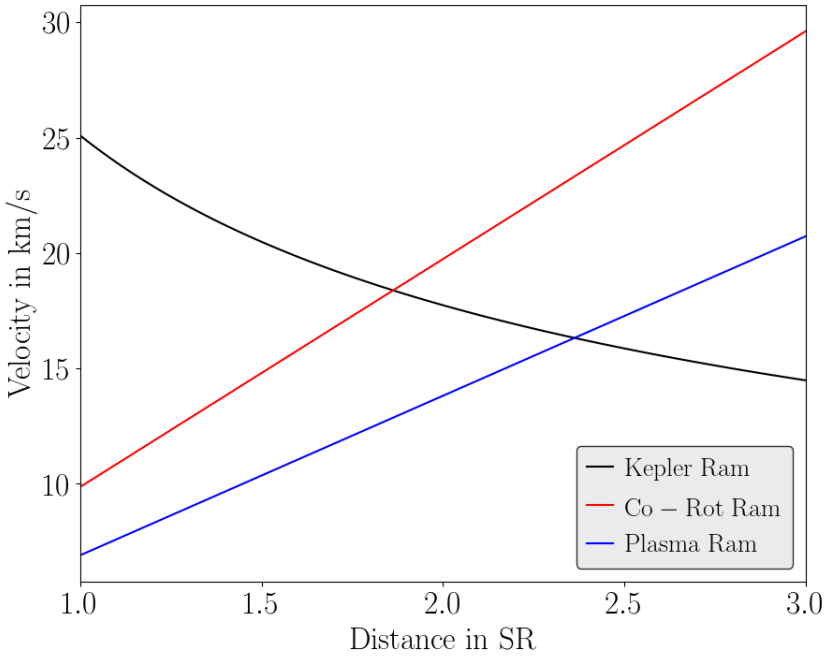


Figure 3.9.: Kepler (Dust) -, Co-Rotational- and Plasma-Ram velocity in km/s vs. the radial distance from Saturn. From Albin [2014].

4. Application: The Meteor Stereo Camera System CILBO

This chapter is about the European Space Agency's (Meteor Research Group) Canary Island Long-Baseline Observatory (CILBO) [Koschny et al., 2013]. The system is a double-stationary meteor observatory located on Tenerife and La Palma, respectively. First, a quick introduction is given about the scientific objectives. Then, a section provides an overview of the system itself, the mechanical, optical and electronic configuration, the data acquisition and reduction pipeline and the trajectory computation tool Meteor Orbit and Trajectory Software (MOTS) [Koschny and Diaz del Rio, 2002].

This thesis focuses on new data analysis methods and technical unbiasing. Thus, a following section describes the technical biases and effects based on calibration data and actual meteor recordings. Although both cameras have the same configuration, a bias is present in the velocity and brightness measurements. Further properties, like the accuracy and proper velocity fitting are described that are all the basis for a proper trajectory computation. The following section describes then a Monte-Carlo based method to compute orbital elements and other astrodynamical properties. The algorithm considers the previously described biases and is a complete extension package for the MOTS routines that is a adjacent orbit computation method beside other methods e.g. developed by Gural [2012] or Dmitriev et al. [2014].

The generated database that contains over 12,000 stereoscopic measurements, is analysed on consistency and accuracy and is a sufficient basis for future scientific research.

4.1. Scientific Objectives

Meteoroids and interplanetary cosmic dust particles are omnipresent in the Solar System. The dust grains are remnants of the formation of the Solar System, are ejected from Comets and active Asteroids or are relicts of catastrophic collisions of minor bodies. In any case, these particles carry miscellaneous information, like astrodynamical properties or the chemical composition of the source. Furthermore, meteoroids are hazardous for (autonomous or human-carrying) spacecrafts [Drolshagen et al., 1995, Graham et al., 2001].

In order to understand the dust and meteoroids population, first, Grün et al. [1985] created a dust model that is still a valid standard for the dust population at 1 Astronomical Unit (AU). Based on Zodiacal Cloud measurements from the COBE DIRBE instrument Dikarev et al.

[2005] derived the Interplanetary Meteoroid Engineering Model for the European Space Agency. Further research were then continued by Soja et al. [2015c], who developed the Interplanetary Meteoroid Environment for Exploration. Soja et al. [2015c] applied the developed model successfully on the comet 67P and did further studies on meteor streams [Soja et al., 2015b].

To contribute scientific measurements and input for the various models and in order to understand the dynamics of the meteor showers and their chemical composition Koschny et al. [2013] defined the following two major tasks of the camera system:

1 To study physical and chemical properties of meteoroids, taking into account the modifications of the meteoroid properties during their flight in the Solar System, to constrain the physical and chemical properties of their parent body.

2 To study the variability of the background dust flux in the Earth environment during a complete year.

As already mentioned, especially the flux studies have been recently extensively investigated. E.g., Drolshagen et al. [2017] filled the meteoroid population gap between the Grün model and the model by Brown et al. [2002] that describes the population of larger Near-earth objects in Earth's vicinity.

4.2. Set-Up

The CILBO system has been partly installed in July and December 2011 on the canary islands and started its continuous and automatic observations a few months later. Currently (December 2017), the system does not operate due to technical issues, probably related to the degeneration of exposed cables to the environment.

CILBO consists of two separate sub-systems with different name affix that are build similarly [Koschny et al., 2013]. Since the system requires a power supply and internet access, one system is located at the Izaña Observatory, the other one at the Observatorio del Roque de los Muchachos on Tenerife and La Palma, respectively. Both stations are installed approximately 120 km apart and have an overlapping Field-of-view (FOV) at an altitude (w.r.t. mean sea level) of 100 km between both islands. Figure 4.1 shows a schematic sketch of the location and geometry of both systems. The red (top left) and green (bottom right) pyramids indicate the FOV.

Table 4.1 summarises the location information of both systems as well as the boresight crossing coordiantes.

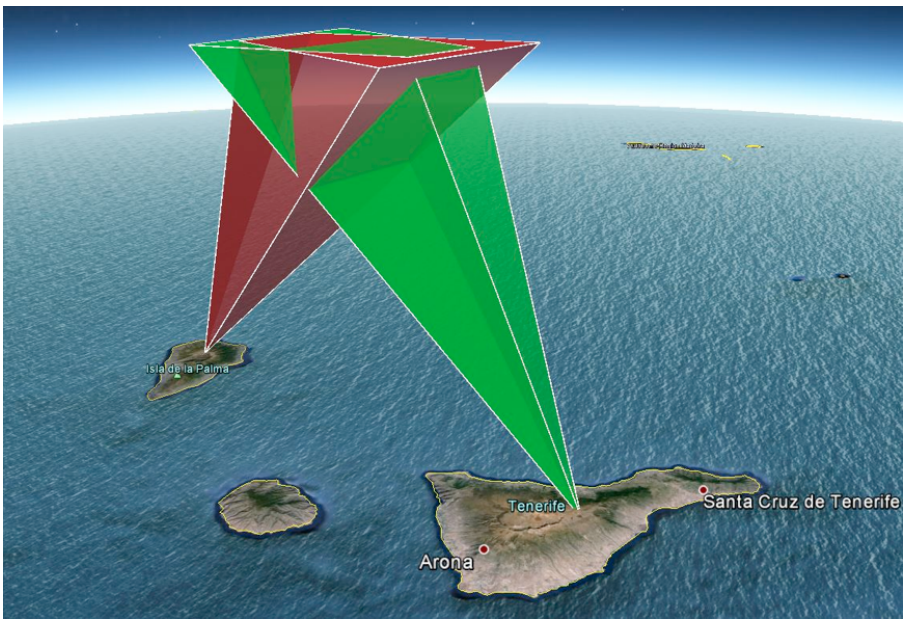


Figure 4.1.: Sketch of the location and Field-of-view (FOV) of both Canary Island Long-Baseline Observatory (CILBO) sub-systems. Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9 are located on Tenerife (bottom right island) and La Palma (top left island), respectively. The green and red pyramids indicate the corresponding Field-of-view (FOV) that overlap on top at an altitude of around 100 km. Image: Google Maps, XML data provided by: D. Koschny (priv. comm. 2015).

Table 4.1.: Location and pointing information of CILBO. The first column indicates the sub-system, the second and third one show the geographical coordinates and the last column shows the elevation in meters. The last row shows the boresight crossing parameters. From: Koschny et al. [2013] (edited).

Sub-System	Geo. Long.	Geo. Lat.	Elevation / m
CILBO-T	28°18′04″ N	16°30′43″ W	2395
CILBO-L	28°45′36″ N	17°52′57″ W	2327
CILBO boresight crossing	28°32′00″ N	17°10′00″ W	100,000

4.2.1. Configuration

In total, both sub-systems consist of 3 cameras called ICC. ICC7 (Tenerife) and ICC9 (La Palma) are used for stereoscopic observations. With these cameras, a meteor's trajectory can be reconstructed that crossed the overlapping observation area and was detected by both cameras. Furthermore, CILBO-T has an additional camera (ICC8) that is similar to the other cameras, but is used for spectroscopic observations of meteors. For this purpose, in front of the camera is a transmission grating mounted with a line density of 661 mm^{-1}. A block diagram of the configuration can be seen in figure 4.2

To fulfil the scientific requirements (stellar magnitude of 6.5 mag and an accuracy better than 1 arcminute) Charged-Couple Device (CCD) cameras are used with a resolution of 576×768 $pixel^2$ with an attached image intensifier. The lens is a Fujinon CF-25L that has a focal length of 25 mm and an aperture up to f/0.85. This camera set-up allows an observation of an

area of $22° \times 28°$ with 25 fps and an 8 bit dynamical range. These 3 cameras are called ICC with an attached version number (here: 7 to 9). Furthermore, the sub-systems are equipped with heaters to prevent dewing.

To protect the equipment from environmental conditions the electronics are installed on a massive mounting in a housing. The complete set-up is stationary and has a concrete foundation as a basis. A roof protects the electronics and covers it during harsh conditions and sun- and moonlight, since a too bright illumination damages the image intensifiers. The roof is controlled by the *Dome control* that is part of the *MRG software suite*. Depending on the date and time the roof is opened for observations. Via a *Power Control*, the ICCs are supplied with necessary power. *Dome Control* uses also streaming data from a weather sensor that is attached to the housing. Depending on e.g., the wind or humidity conditions, the *Dome Control* software closes the roof to prevent damages.

Data from the photometric cameras are send via a Phase Alternating Line signal to the software MetRec [Molau, 1999][1]; are then archived and send to a central server (*ftp server*), where all data from both stations are collected for further analysis. The spectroscopic camera is controlled via *Spcrec*, but is not considered in this work. A detailed description of the data pipeline is given in section 4.2.2.

4.2.2. Data Pipeline

Since this work focuses only on the bias effects and derived scientific parameters from the photo- / astrometric measurements, this section only describes the data pipeline that includes the detection software MetRec [Molau, 1999]. A complete and detailed description of the shown pipeline can be found in Koschny et al. [2013] and Albin et al. [2017].

Both photo- / astrometric ICCs capture continuously images of the sky. The generated Phase Alternating Line signals are streaming into a frame grabber card, namely a *Matrox Meteor II*. The streamed signal is then digitised and analysed by MetRec. First, the signal is down-sampled to 384×288 pixel^2, and certain frame regions are masked, like e.g., a superimposed data or time frame in the corner of the video. The images are then differentiated. For example: The celestial equator moves approximately a quarter of an arcminute per second. With a wide-field set-up like CILBO the pixel-dwell time of a star is larger than this movement (25 fps sampling rate). Thus, the differentiation leads to a removal of slowly moving objects. The resulting images are then divided by a flat-field, to correct the illumination cut-off of the lens. During an observation night, several flat-fields are generated based on the continuous video stream. The resulting light images (images for further scientific / analysis computations) are then used to detect meteors.

The meteor detection algorithm concept is based on the ideas of Molau and Nitschke [1996]. Basically, the algorithm sums the values of neighbouring pixels in different directions. Based on

[1] `http://www.metrec.org/`, 07.12.2017

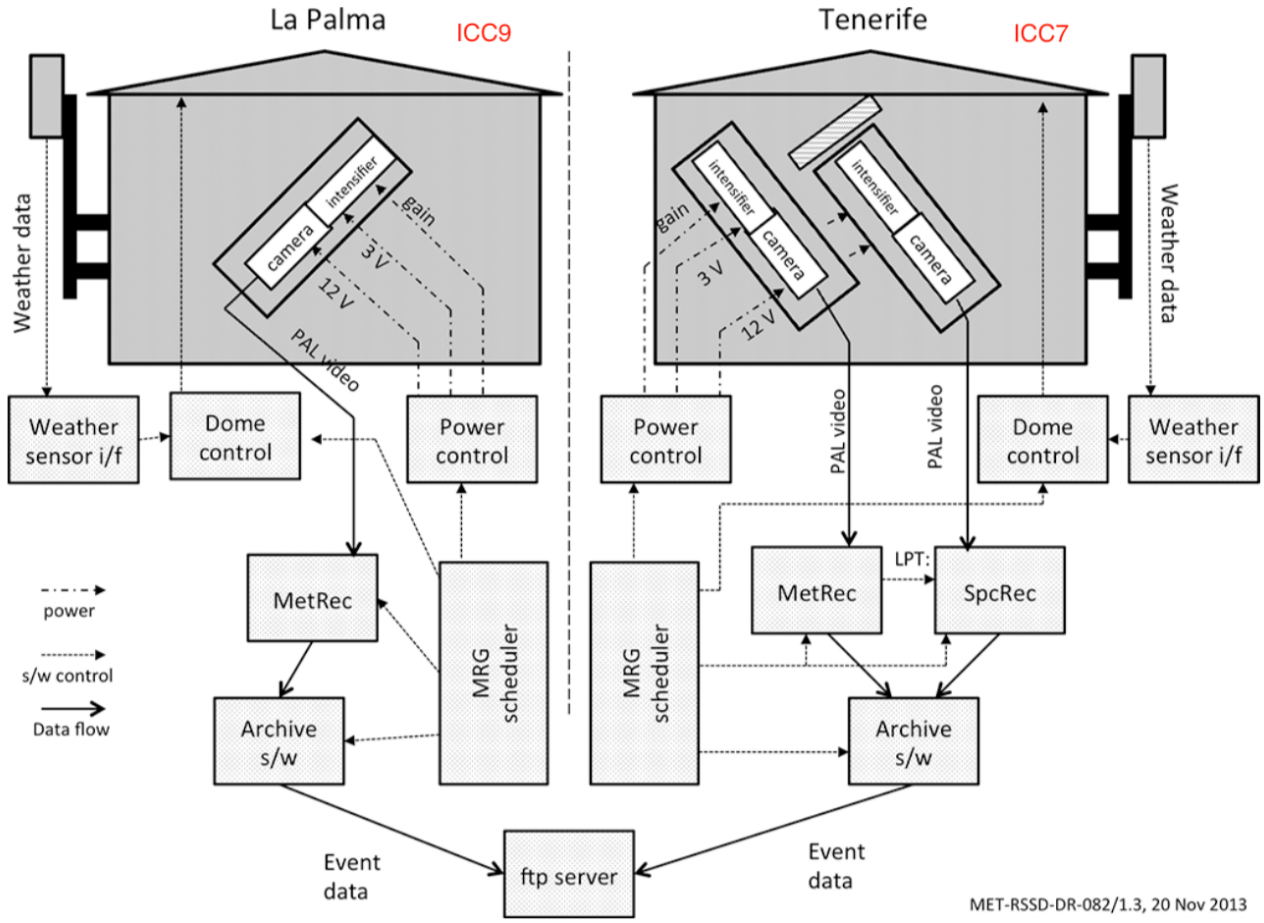


Figure 4.2.: Block diagram of the Canary Island Long-Baseline Observatory (CILBO) system. The diagram shows the architecture of both stations, since Canary Island Long-Baseline Observatory (CILBO)-T has an additional Intensified CCD Camera (ICC) for spectral measurements. The dashed line between both stations indicate the separation of both sub-systems. A detailed explanation can be found in the main text. From: Koschny et al. [2013].

a detection threshold, the program checks whether bright "regions of interests" can be identified. Since a meteor's trajectory follows a great circle on the sky, this movement is considered, too. If a meteor is detected, several parameters are derived, like the number of frames, brightness information and the photometric centres of the meteor on each frame. Additionally it stores the images and stacks them to a movie to allow a visual analysis and confirmation of the meteor. To ensure a complete trajectory capture, MetRec stores also 3 frames before and after the first and last confirmed detection, respectively. An exemplarily image of a meteor is shown in figure 4.3. The meteor was recorded by ICC7 of the station CILBO-T on 05.10.2013 between 05:13:58.38 UTC and 05:13.59.09 UTC. The meteor had two brightness peaks and was observed simultaneously by ICC9.

4.2.3. File structure

All meteor event concerning data are stored in a so-called information file with the ending *.inf* (hereafter called inf-files) that have the name structure HHMMSS.inf, with HH as hours,

Figure 4.3.: Stacked image of a meteor that was detected and recorded by Canary Island Long-Baseline Observatory (CILBO)-T (Intensified CCD Camera (ICC)7). The meteor was recorded on 05.02.2013 between 05:13:58.38 UTC and 05:13.59.09 UTC.

MM as minutes and SS as seconds. An example is given in listing 4.1, where the content of a detected meteor is shown that was observed at 00:36:23 UTC on 16.02.2016. The date and time are shown in the first 2 header lines. Header line 3 lists the path to a so-called reference file, where the used reference stars are listed for the astrometric computations. Line 4 gives the total number of frame counts (including the +/- 3 frames around the actual observation). Line 5 shows the column names of the following table. The first column is a consecutive index number starting at 01. The time in seconds is given in the second column. Columns 3 to 7 list the brightness in magnitude and photometric centre of the meteor in CCD X-Y-coordinates and equatorial coordinates (alpha: α, delta: δ), respectively. For 6 frames a brightness value (in mag) has been determined. However, for all frames a photometric centre has been computed, since MetRec extrapolates the centres. Line 8 and 11 list the X-Y and equatorial coordinates that are corrected by a 3rd-order polynomial fit to compensate the distortion. For the CILBO set-up, this correction is only applied on actual measured equatorial coordinates. The very last column indicates, whether a frame was used to determine the corrected values.

Listing 4.1: Content of 0036023.inf

```
AppearanceDate 16.02.2016
AppearanceTime 00:36:23
ReferenceStars c:\cilbo\metrec\config\20150214.ref
FrameCount 12
#   time  bright    x      y      alpha     delta    c_x    c_y    c_alpha   c_delta   use
01  23.711 -----  0.817  0.411    6.0752   43.528  -----  -----  --------  --------  no
02  23.751 -----  0.824  0.404    6.0557   43.704  -----  -----  --------  --------  no
```

```
03  23.791 ------  0.831  0.397    6.0361   43.880 ------ ------ -------- -------- no
04  23.831   4.1  0.839  0.390    6.0166   44.054 ------ ------   6.0157   44.062 yes
05  23.870   2.0  0.846  0.384    5.9971   44.224 ------ ------   5.9987   44.218 yes
06  23.910   1.4  0.852  0.378    5.9795   44.382 ------ ------   5.9797   44.378 yes
07  23.950   1.8  0.859  0.371    5.9609   44.539 ------ ------   5.9615   44.538 yes
08  23.990   3.1  0.866  0.365    5.9434   44.697 ------ ------   5.9430   44.696 yes
09  24.030   4.0  0.872  0.359    5.9248   44.851 ------ ------   5.9245   44.855 yes
10  24.070 ------  0.879  0.353    5.9063   45.005 ------ ------ -------- -------- no
11  24.110 ------  0.885  0.347    5.8877   45.159 ------ ------ -------- -------- no
12  24.150 ------  0.892  0.340    5.8691   45.312 ------ ------ -------- -------- no
```

Additionally to the inf-files, MetRec generates log-files (*.log*) every night, named YYYYMMDD.log, with YYYY as year, MM as month and DD as day (observation beginning in the evening). These files contain miscellaneous data, like the set parameters for the software, the total observation time, astrometric quality information and some statistics of the total observation session. This includes the flux density, number of detected meteors as well as a possible classification of the meteors (sporadic as well as shower meteor). For each meteor detection, the log file stores meta information and the mean astrometric accuracy. After the record, the recognition algorithm is set again to stand by. Note: If a second meteor appears during the record of another, previously appearing one, the second meteor overwrites the data in the inf-file, however, two entries can be found in the log-file. Since a log-file can be vast, listing 4.2 shows a small fraction of the log-file 20160215.log, where the corresponding entry of the inf-file 4.1 is shown. Each line of a detection starts with a time stamp in UTC. First, a consecutive number (here 14) is assigned to the meteor. The starting and ending point of the meteor is shown in CCD coordinates. Then the number of frames, duration of the observations, direction and velocity in pixels/second is listed. The Signal-To-Noise ratio (SNR), shower classification, brightness, equatorial information as well as the astrometric accuracy allow a quick overview of the recorded event.

Afterwards, a few data are stored, a health check is performed and the recognition software restarts. The complete analysis and storing takes around 1 s to 2 s.

Listing 4.2: Content of 20160215.log

```
00:36:23 Meteor #15 at (0.839,0.390)->(0.872,0.359) frames=6 dur=0.22s pixel=22 dir=317ø vel
    =6.3ø/s snr=7.4 shower=SPO bright=1.4mag (6.016h,44.06ø) -> (5.925h,44.85ø) acc=0.7'
00:36:24 Saving meteor data of #15 ... ok!
00:36:24 Saving image band of 12 frames of meteor #15 ... ok!
00:36:24 Saving sum image of meteor #15 made of 12 frames ... ok!
00:36:24 Restart recognition
```

After the end of an observation night, all generated files are send and stored on a centralised FTP-Server. Each sub-system generates a directory separately, where the inf- and log-files are stored as well as the (stacked) images and additional meta data. A complete list can be found in Koschny et al. [2013, p.345]

4.2.4. MOTS

One major scientific task in meteor research is the analysis of astrodynamical parameters of the meteors, to e.g., identify the sporadic background sources, to classify streams or to determine possible parent bodies. The provided data on the FTP-Server contain astrometric information from both stations. With these data, the known date and time of the observation and the geographical coordinates and altitude of both camera stations, the trajectory of simultaneously detected meteors can be reconstructed. For this purpose, Koschny and Diaz del Rio [2002] developed the PYTHON2 based software MOTS.

The computational algorithm of MOTS is visualised in figure 4.4. The sketch shows schematically the curvature of the Earth and the position of two meteor stations named *S1* and *S2*, respectively. The path of an entering meteor is illustrated as a thick solid line between the beginning B and end E point of the meteor's appearance.

Both cameras record the meteor individually (different number of frames, movement on CCD, etc.). Each determined photometric centre (given in RA / DEC) can be used to compute an according vector w.r.t to the observing station. Since no distance information can be derived from the astrometry, the vectors are theoretically infinite long. With the vectors between *S1* and the photometric centres of the recorded frames, a plane can be computed that is described by the normal vector $\vec{n}$. A plane is mathematically determined by the cross product of two vectors. Here, the RA / DEC corresponding vectors of two different video frames are used. The number of possible cross products is,

$$\binom{v}{2}$$

where v is the number of vectors (corresponding to the number of video frames). An average plane of all cases is then used for further computations.

E.g., with a meteor that was recorded on 5 frames 10 normal plane vectors can be computed. Based on these vectors, an average is computed. Additionally Koschny and Diaz del Rio [2002] added a weighting factor for each "vector pair". Vectors with larger angular distances are less erratic than closer vectors.

The plane that is defined by $\vec{n}$ can be seen in the sketch below, too. A projection of the plane is drawn on the Earth's sphere for better visualisation. To determine the positional vector of the meteor, data from *S2* are now used. For each video frame, the RA / DEC values are computed, too. For each vector the intersection point with the plane is computed. The intersecting point determines the distance between the meteor and *S2*. A corresponding distance and position w.r.t. *S1* can be computed, too. This is visualised schematically with the vector $\overrightarrow{M2}$. The same computational procedure can also be done vis-versa.

With the positional vectors of the meteor and the time stamps of each recorded time frame, a corresponding velocity can be calculated. Here, the velocity can be computed by using two

consecutive vectors and times, or the average between the first and last recorded frame. Thus, each positional vector (but the last one) gets additional velocity information. With a set of state vectors, additional scientific analysis can be performed.

All generated data from the MOTS routine are stored in so-called Detailed Altitude Files. The file format is conveniently similar to METREC's inf-file structure. A detailed header provides information like e.g., the observation date and time, shower classification and astrometric accuracy, obtained from the log-file. A table contains then again an index number, brightness values and the CCD-coordinates. The altitude in meters is given as well as the corresponding error and the deviation between the fitted trajectory line and the determined meteor position. With the given longitudinal and latitudinal information (in degrees), a trajectory can be plotted. Additionally, the distance to the camera station is provided and finally the velocity values. Figure 4.5 illustrates exemplary the positional and brightness information of a simultaneously detected meteor at 05:13:07 Coordinated Universal Time (UTC) on 08.05.2013. The Y and X-axis show the relative CCD coordinates. Thus, the complete CCD ranges from 0 to 1. Here, only a fraction is shown, where the meteor appeared. Two lines indicate the detected meteor on ICC7 and ICC9. Each dot represents the photometric centre and a grey scale displays the corresponding brightness in magnitudes. In the shown case, the La Palma and Tenerife camera recorded 9, respectively 5 video frames with a peak brightness around 2.5 mag.

MOTS contains several classes and functions for further data parsing. For example, the generated Detailed Altitude Files can be converted to an XLM file that can be read and visualised by GOOGLE EARTH.

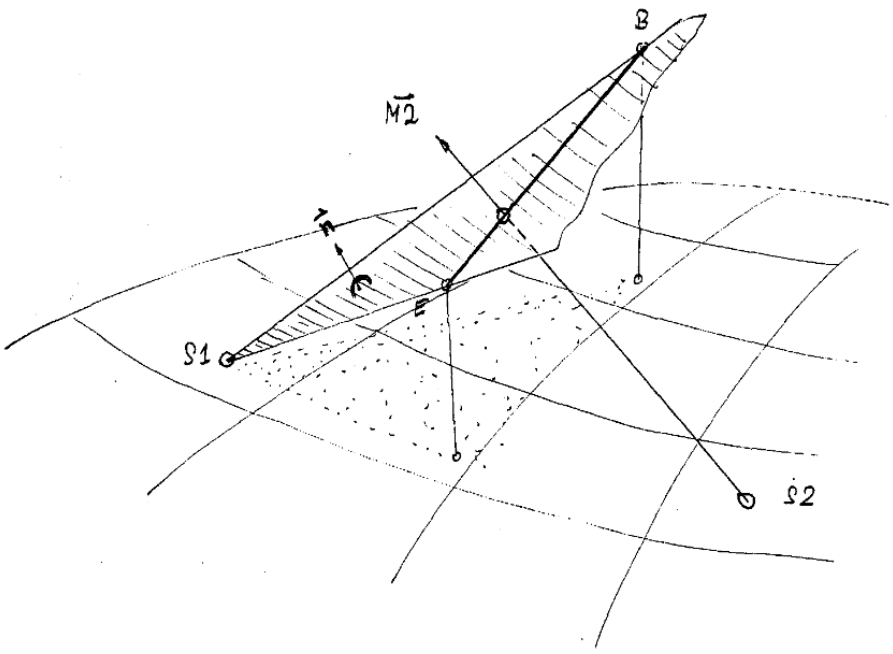


Figure 4.4.: Graphical representation of the Meteor Orbit and Trajectory Software (MOTS) algorithm. *S1* and *S2* are two meteor observing stations that recorded the path of a meteor (thick solid line between B and E). The RA, DEC corresponding vectors w.r.t. *S1* lead to a plane with the normal vector $\vec{n}$. With the RA / DEC information of a meteor from *S2* a vector can be determined $\overrightarrow{M2}$. A complete description and explanation can be found in the main text. From: Koschny and Diaz del Rio [2002].

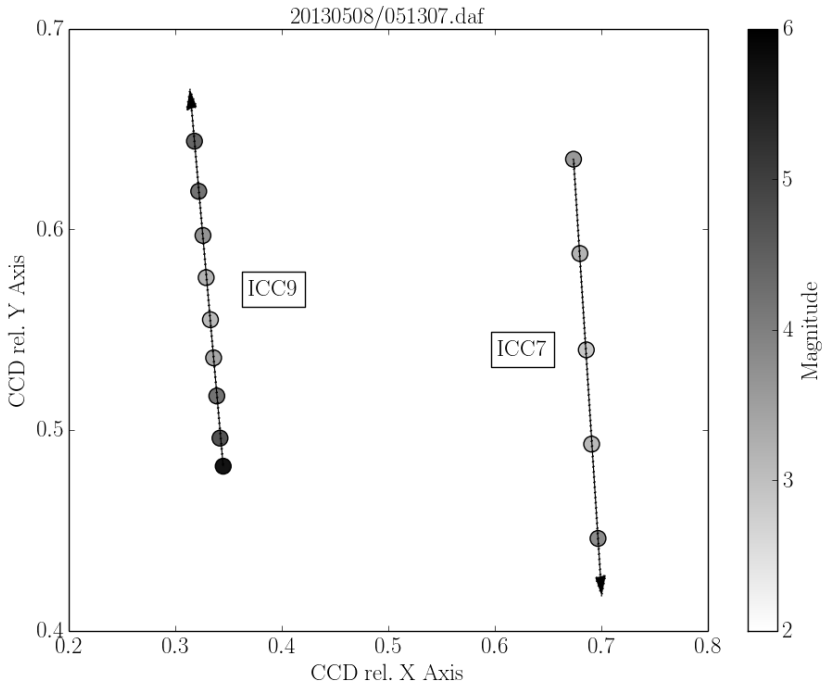


Figure 4.5.: Example of a simultaneously detected meteor at 05:13:07 UTC on 08.05.2013. The data were extracted from a Detailed Altitude File. The dots represent the determined photometric centres from ICC7 and ICC9 and a dashed arrow indicates the flight direction, however has been artificially added for better readability and is not stored in the Detailed Altitude File. Relative (normalized) Charged-Couple Device (CCD) coordinates show the positional information. A grey scale encodes the brightness information. From Albin et al. [2015a].

4.3. Statistical properties of observations

To achieve the scientific goals of CILBO the data need to be analysed depending on the objective, like meteor flux, brightness respectively size and velocity distribution, flight direction of the dust particles or the computation and analysis of astrodynamical properties. Each task requires miscellaneous data from the CILBO system, stored in the different formats. All astronomical setups or observational data measurements have a mechanical, optical, electronic and algorithmic part. Each part, depending on the chosen parameters and properties has certain advantages and disadvantages to fulfil the objectives. This leads to biases in the detection chain of meteors, leading to e.g., a shift in the brightness or velocity distribution; or leads to a biased detection performance, although identical system are used.

One major selection bias effect in astronomy is the *Malmquist-Bias* that describes a selection bias to brighter stars. Stars that are intrinsically fainter are under-represented, since the apparent magnitude of an objects scales with the distance to the observer. The detectability of these objects decreases. Without an unbiasing, brighter objects would be "over-represented" in the statistics. Gunnar Malmquist described this effect first in 1922 [Malmquist, 1922] and future work lead to an un-biasing of star populations of our stellar neighbourhood. One can

expect similar bias effects in remote meteor observations. However, the angular / apparent velocity of a meteor adds to its brightness. This leads to a higher dimensional bias problem that is shown here.

This section describes bias effects that affect the detection of meteors with CILBO. However, the implications from this analysis can be applied on other stereoscopic meteor camera systems, too. First attempts of the analysis and bias understanding have been published in Albin et al. [2015b] and Albin et al. [2015a]. Later in Albin et al. [2017] a detailed overview was given and is partly shown in this section.

For the technical bias analysis data was available from 13 September 2011 to 31 August 2015 including 51,062 inf-files for ICC7 and 56,951 inf-files for ICC9, respectively. Additionally, in the same time period a total of 925 (ICC7) and 913 (ICC9) log-files are used. So, CILBO was operated at around 65 % of the nights in the time frame (offline time: maintenance, technical issues, etc.).

4.3.1. Error / Bias overview

The three most prominent sources of bias effects are described in this section in each paragraph. First, effects from the instrumental setup are shown (mechanical, electronics, pointing direction), then the expected algorithmic and finally statistical errors are listed. Albin et al. [2017] provides a sophisticated list of possible errors, based on 1.5 years of research.

Instrumental errors / biases

1 Mechanics: The mounting, cameras and lenses are affected by varying temperatures. Thermal expansions and contractions of the set-up lead to a temperature depending change of the precise pointing position. This can affect the exact position measurement of the meteor.

2 Optics: CILBO uses a wide-field lens with 22×28 deg^2. This lens parameters lead to an illumination / light transmission drop with larger distances from the optical axis, due to vignetting and the tangent effect. (Possibly, additional cut-off effects might be caused by the image intensifier, too).

3 CCD / Electronics: The spherical coordinates of the meteor are projected on a flat CCD chip. This causes a distortion that needs to be corrected. The effect is linked with the algorithmic error #3.

4 CCD / Electronics: Optical sensors have miscellaneous noise sources that affect the astro- and photometric measurement like: dead / hot pixel, thermal noise, readout noise. Furthermore, the image intensifier causes also random noise. All effects need to be corrected by dark and bias frames; these frames vary in time.

5 CCD / Electronics: The signal of the CCD is interpolated, since the video stream pipeline is a Phase Alternating Line signal that does not match the pixel resolution of the chip.

6 CCD / Electronics: The CCD has an interline-transfer readout. The original idea of a CCD is to read the pixels out sequentially (line by line). For each read-out cycle, the chip in the ICCs masks every second line and uses these lines to transfer and buffer the information from the illuminated, neighboring lines. This affects the position accuracy of each recorded meteor (25 frames per second).

7 Electronics: Timing errors / shifts. Astrometrically accurate measurements require a precise time determination. For this purpose CILBO synchronizes regularly the internal clock with a time server. However, even a small shift of only 1 second leads to a positional error in right ascension of a quarter arcminute.

Algorithmic errors / biases

1 Wake: Meteors may develop a wake and train during their entry in Earth's atmosphere. Ion trails, fragmentations, etc. cause a "stretched" glowing of the meteor, moving the maximum brightness, and thus photometric center behind the actual position of the meteor. Wakes were first described with analogue cameras on Super-Schmidt telescopes [McCrosky, 1958]. Figure 4.6 shows exemplary wakes / trains of a meteor from McCrosky [1958]. Figure 4.7 illustrates schematically the cause of the effect. Wakes / trains develop towards the end of a meteor's appearance. The photometric centre shift may cause a reduced velocity determination of the meteor.

2 Blooming: Very bright meteors cause an overexposure of the CCD. Electrons "overflow" to neighboring pixels. Consequently, the photometric centre cannot be determined correctly and might also affect the velocity calculation.

3 Distortion: As described in the *Instrumental errors / biases* item 3, the spherical coordinates are projected on a flat CCD chip. To unfold and consider this effect, MetRec uses a 3rd-order polynomial fit. This however might not be sufficient enough and cause a systematic shift between the computed and actual position of the meteor.

4 Position / Velocity fit: To determine the position of a meteor a linear or quadratic fit function is used that results in a constant and linear velocity equation, respectively. This procedure might not be sufficient enough, especially for long meteor appearance (corresponding to a larger number of recorded video frames).

5 First / last frame: CILBO records the sky continuously with 25 frames per second. Thus, the video recording is not triggered by a meteor's appearance. Consequently, in the first and last video frame of a detection, a meteor appears and disappears respectively, at

a random time. The exact times cannot be determined, causing a bias in the position determination of the first and last photometric centre.

6 Quantization: The position data are stored in relative CCD coordinates, varying between 0 and 1 with a precision of 3 decimal places. This corresponds to an error of around 0.3 pixel. If the meteor's position is re-computed based on these values, the quantization leads to a systematic position and velocity error. It is expected that this effect is less dominant for meteors recorded on more frames.

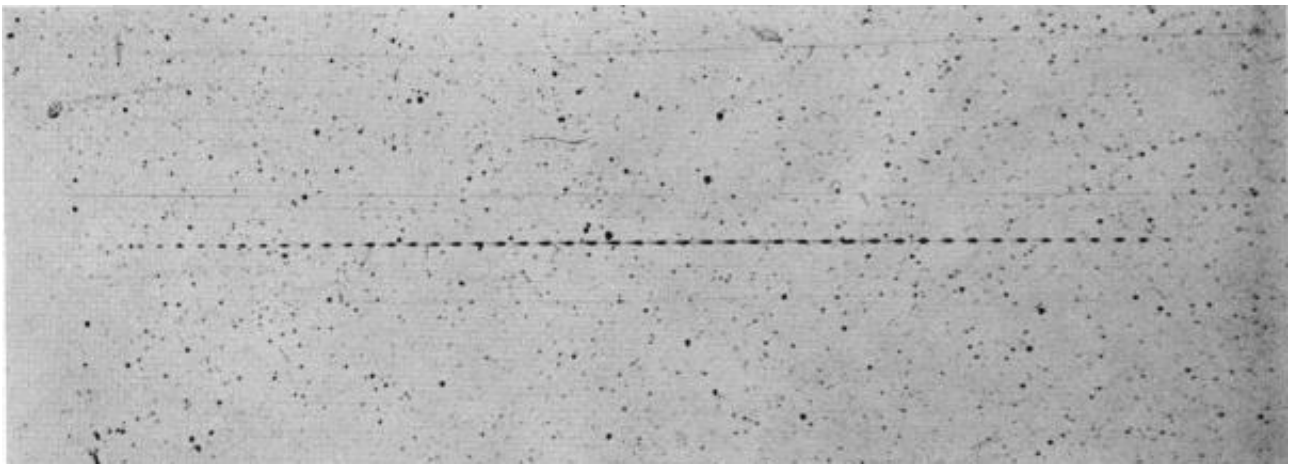


Figure 4.6.: Recording of a meteor with an entry speed of 26 km/s. The shutter breaks cause a "splitting" of the meteor's appearance, to derive e.g. the velocity. From McCrosky [1958].

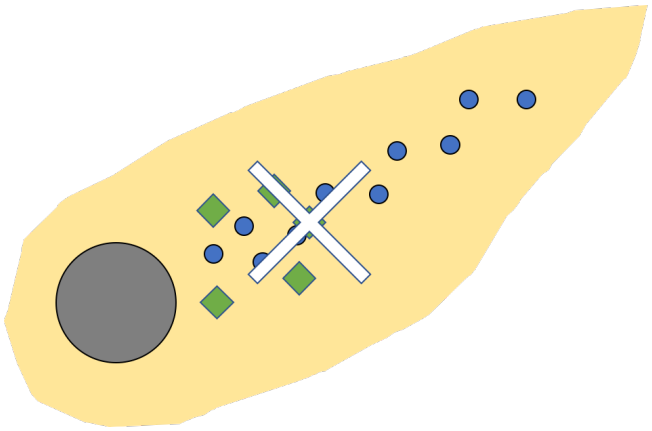


Figure 4.7.: Schematic illustration of a meteor wake. The grey area indicates the entering meteoroid that leaves fragments of particles (diamond shaped) and ions (blue shaped) behind. Thus, the resulting illuminated area (light orange / yellow around the meteor) is widen. An automatic detection software might set the centre of the illumination area as the photometric coordinates of the centre, thus, shifting it to the end of a meteor (indicated with a white cross). Detailed models and analysis of wake formations can be read in a recent work by Stoken [2014].

Statistical errors / biases

1 Statistical random errors / noise appear in any experiment and are due to the probabilistic nature itself. It affects the position (consequently velocity) and brightness determination

and is depending on miscellaneous parameters, like accuracy, number of frames, varying electronic noise, brightness determination etc.

4.3.2. ICC sensitivity / Detection efficiency

The analysis of CILBO's observational and detection performance starts with a fundamental instrumental calibration image, the flat-field and its implications in meteor astronomy. Telescopic observations start or finish during dusk and dawn, respectively to determine the nightly flat-field. The less bright, homogeneously illuminated clear sky is used as a flat-field calibration image. A sequence of images is then taken and a (normalised) median image is computed. Light images are then (after subtracting the darks) divided by the flat-field image. To prevent any appearing stars in the image, the star tracking systems are turned off. Alternatively, for small telescopes a so called flat-field box can be used; a homogeneously diffuse glass plate that is back-illuminated by several lamps (e.g., LEDs).

CILBO is a steadily mounted instrument that does not track stars. However, since a sensitive image-intensifier is used, neither a flat-field box nor the dusk or dawn sky can be used for calibration purposes due to the to high brightness. MetRec generates flat-field images during the night based on the night sky, but the low brightness and resulting noise is not feasible for a sophisticated analysis.

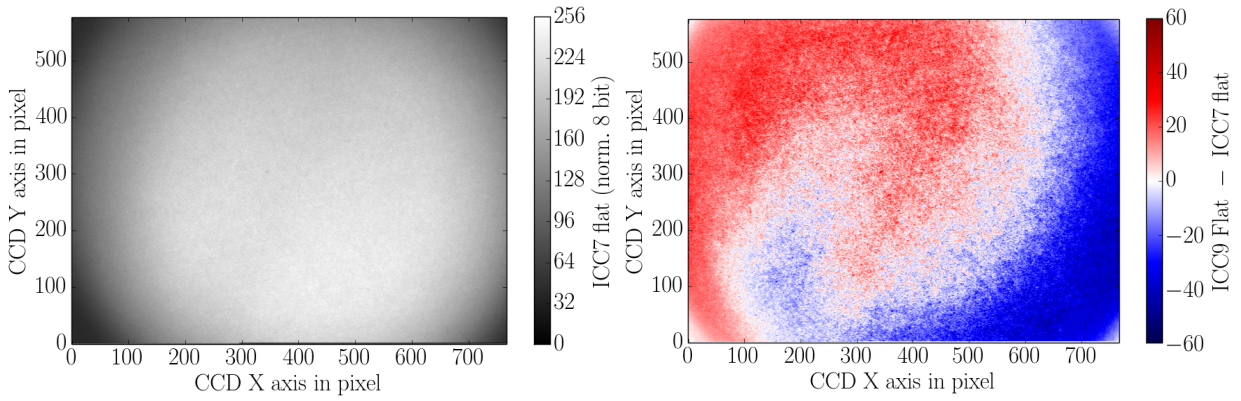
Thus, Albin et al. [2017] used around 10 images for each ICC system during foggy weather conditions at night[2]. CILBO's weather monitoring software closes the dome and stops any observational operation if the humidity reaches a certain threshold. For this purpose the automatic routines have been overwritten by a manual operation.

Figure 4.8 a) shows a resulting flat-field image of the ICC7 system. The image is a median image and represents the brightness values in 8-bit for each individual pixel position. The X- and Y-axis of the CCD chip are not down-sampled and the corresponding brightness values are colour coded. Apparently, the optical system of ICC7 is slightly off-centre to the right w.r.t to the optical axis due to an offset in the optical system. ICC9 has a similar flat-field. Both system show a strong decrease to the edges and especially to the corners of the CCD due to the vignetting and projection-cut-off effects.

Since both flat images appear similarly, figure 4.8 b) is shown to visualise the differences between both flat fields. Here, both 8 bit median images are subtracted: ICC9 - ICC7. A color code indicates the deviations. Redder or bluer colours indicate a higher sensitivity for ICC9 and ICC7 w.r.t. to the other system, respectively. A white colouring indicates no deviations between both systems. Both systems have a shifted or off-centred optical system.

It is expected that several observational parameters are affected by the vignetting / cut-off of the ICC systems. An ideal or perfect observational system with no vignetting or projection effects would have the same normalised brightness value all over the camera chip. A computed

[2]As mentioned, this paper is a comprehensive paper of the bias effects that are shown here.



a) ICC7 b) ICC9 - ICC7

Figure 4.8.: a) Flat image of the Intensified CCD Camera (ICC)7 system. The plot shows an 8 bit median of around 10 images. The X- And Y-axis show the pixel position of the CCD chip (not down-sampled). A colour bar on the right indicates the median value. From Albin et al. [2017]. b) Difference image of both flats: Intensified CCD Camera (ICC)9 - Intensified CCD Camera (ICC)7. A colour code represents the deviation between both 8 bit medians and indicates the tilted or off-centre mounting of the optics.

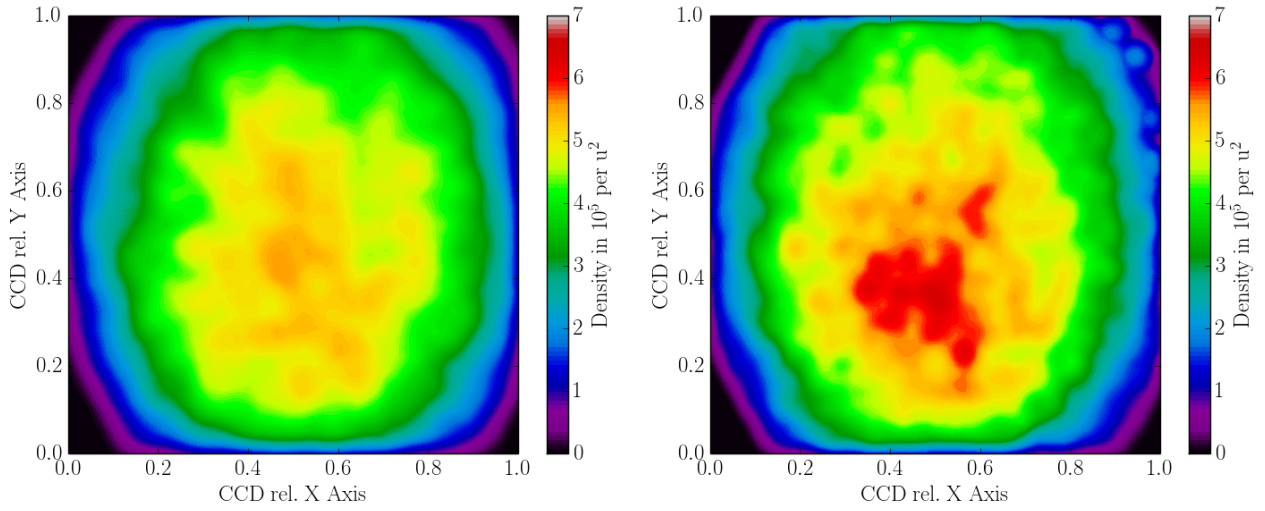
flat-field by MetRec can compensate the light image, however a flat-field cannot compensate the decreased SNR to the corners. Thus, the detectability of meteors is affected negatively to the edges and corners.

To verify this effect, an analysis is performed on actual measurements. Figure 4.9 a) and b) show the 2-dimensional detection distribution of meteors for ICC7 and ICC9, respectively. For the analysis, all X and Y photometric positions are taken into account from the inf-files. Since brighter, respectively darker appearances are rather rare and / or difficult to detect properly, only observations of meteors are taken into account with a brightness between +1 and +6 mag. The CCD coordinates are normalised (0 to 1) and with a Kernel Density Estimator (KDE) algorithm an Epanechnikov kernel[3] is applied on the data. The according bandwidth is determined using a cross validation algorithm for both cases. A color code bar on the right side of each figure shows the resulting density in u^2. 1 u^2 is introduced as the normalised area of the complete CCD; the letter u is artificially set and does not abbreviate any technical or physical quantity.

As expected from figure 4.8 a), the detectability decreases to the edges and corners, respectively. ICC9 shows a significant detection maximum off-centred from the optical axis to the bottom. The vignetting and projection effects result in a reduced detectability by a factor of 7 between the corners and the optical axis. If not considered, consequently this biases scientific results, like e.g., meteor flux computations.

To describe the bias effects quantitatively, two meteor properties are considered, the determined brightness and velocity on the CCD chip as shown in Albin et al. [2015a] and Albin et al. [2017]. Furthermore, a new quantitative parameter is introduced, namely the *signal ratio* [Albin

[3]The Epanechnikov kernel proportional to $1 - t^2/h^2$, where t is the variable and h is the bandwidth. For $t > h$ the kernel results in 0.



a) ICC7 b) ICC9

Figure 4.9.: 2-dimensional density plot of determined photometric centers. Only those data are considered with a brightness between +1 and +6 mag. The X and Y axis show the Charged-Couple Device (CCD) chip coordinates, normalised from 0 to 1. The density is colour coded and given in total numbers per u^2, where 1 u^2 is equal to the total chip area. Both figures are computed based on an Epanechnikov kernel (Kernel Density Estimator (KDE)). a) and b) represent the data for Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9, respectively. From Albin et al. [2015a].

et al., 2015a] or *detection efficiency* [Albin et al., 2017], respectively. The detection efficiency is defined based on the following assumptions: A meteor (either from a stream or sporadic source) occurs randomly on the sky and thus randomly on the CCD chip. An uniformly distributed density plot would be expected in an ideal case (no vignetting, projection cut-off etc.). The shown deviation from an uniformly distributed density results in a detection loss. To estimate the detection loss one assumes that around the optical axis (for CILBO slightly offset) the camera system detects all possibly detectable meteors within the technical range of the CILBO system. The maximum value in the density plots is then extrapolated over the whole CCD chip for both ICCs. The ratio between the extrapolated uniform distribution and the actual one is defined as the *signal ratio / detection efficiency*.

The overall *detection efficiency* d_{eff} for both systems, based on figure 4.9 a) and b) is:

$$d_{\text{eff, ICC7}} = 0.609 \pm 0.002$$
$$d_{\text{eff, ICC9}} = 0.565 \pm 0.002$$

So, the detection loss of both cameras is between 40 % and 45 % (in the brightness range of +1 mag to +6 mag).

To get a more sophisticated analysis, first, the *Detection Efficiency* is analysed vs. the brightness of a meteor's appearance. For this purpose, the photometric and brightness data from the inf-files are separated in 0.2 mag bins. For each brightness bin, the *Detection Efficiency* is computed with the shown KDE method. The resulting values are shown in figure 4.10 for

ICC9. The data for ICC7 however are not sufficient enough for a separate interpretation, since the data are more scattered. To determine an approximate curve progression, a robust Gaussian Process (GP) is applied on both data sets. A black line indicates the mean and the grey area shows the corresponding 95 % confidence interval. For both stations, the function is approximately constant between +2 mag and +4.5 mag. Due to the lack of sufficient data for very faint and bright meteors, the confidence intervals diverge in both cases. However, considering the data points and mean function of ICC9, one sees an increasing and decreasing *Detection Efficiency* for brighter and fainter meteors, respectively. Bright meteors can be detected in the centre as well as at the edges and corners of the CCD, while fainter meteors become only detectable near the centre (optical axis). This results in a brightness depending *Detection Efficiency* and consequently underestimation of the influx of fainter, respectively smaller meteors. For example, the *Detection Efficiency* of meteors in the 5.8 mag bin is around 40 %; around 15 % to 20 % less than the overall value shown above.

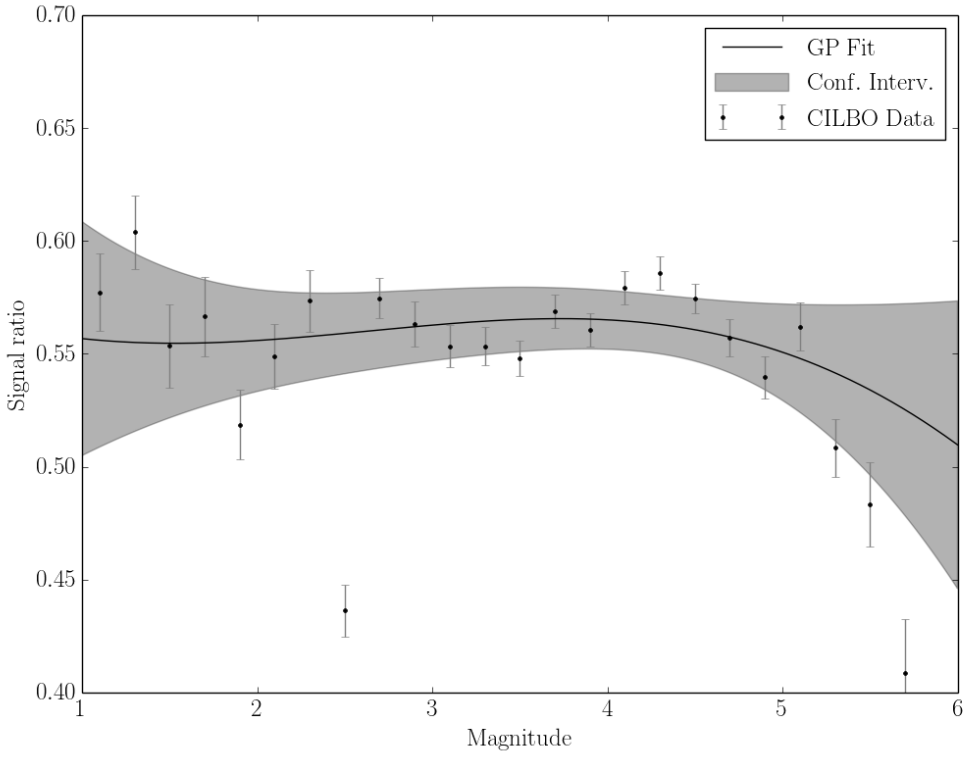


Figure 4.10.: *Signal Ratio* respectively *Detection Efficiency* of Intensified CCD Camera (ICC)9 vs. the brightness of a photometric centre in mag. Each dot with its error bar indicates the result in a 0.2 mag bin. A black line and a grey shaded area indicate a GP fit with the corresponding 95 % confidence interval. From Albin et al. [2015a].

Figure 4.11 shows the *Detection Efficiency* vs. a meteor's velocity. The velocity is given in pixels per second and indicates the apparent movement of the meteor on the CCD chip. Circles and triangles indicate the data for ICC7 and ICC9, respectively. Each dot represents a data bin with a width of 25 pixel s^{-1}. Furthermore, for this analysis, a filtered data set is used that does not consider bright meteors. Due to blooming and resulting errors in the determination

of the photometric centres, the velocity would be biased. Thus, the *Detection Efficiency* values for all velocity bins is less, than the overall value.

For velocities between 0 pixel s^{-1} and around 200 pixel s^{-1} the *Detection Efficiency* remains approximately constant between 40 % and 50 %. Then, the efficiency drops and decreases linearly for higher speeds. Due to the lack of high speed meteor data, the error increases, however one can see that the mean value for both cameras at 380 pixel s^{-1} drops below 20 %. Thus, only every fifth FOV crossing, high speed meteor is detectable with the ICCs.

The velocity depending *Detection Efficiency* results from the duration a meteor staying within a pixel. This pixel dwell time is anti-proportional to the apparent meteor velocity on the CCD chip. High velocities and consequently small pixel dwell times result in a worse SNR than slower meteors. This is due to the less light exposure on individual pixels, and consequently less generated charge (electrons) in these pixels. Since the sensitivity of the CCD drops to the edges and corners, the effective detection area for fast meteors is less prominent, than for slower ones.

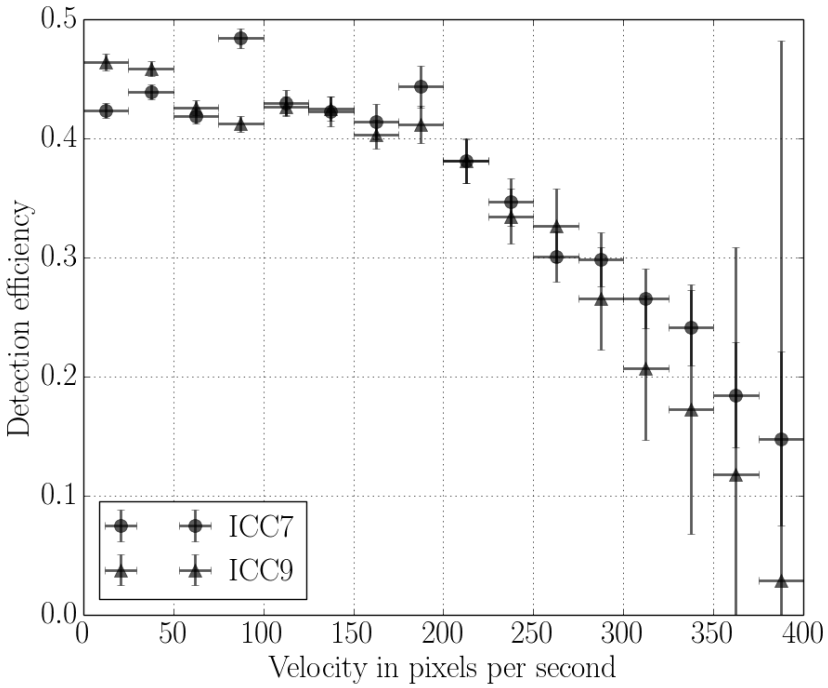


Figure 4.11.: *Signal Ratio* respectively *Detection Efficiency* vs. the down-sampled velocity of a meteor. The velocity is given in pixels per second. Circles and triangles indicate the data for Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9, respectively. From Albin et al. [2017].

As a rule-of-thumb, the computed flux density of meteors, recorded with CILBO needs to be corrected at least by a factor of 2. Also the bias of fainter and slower meteors needs to be considered, since an underestimation of fainter meteors biases the population index r of the shower (or sporadic source) leading to a biased computation of the stream age.

Furthermore, due to the steady installation of the CILBO cameras the angular distance between a shower's radiant and the cameras' boresights varies during the night. Consequently

the apparent velocity changes, too (close distance between camera and radiant leads to a slower movement of the meteor on the CCD chip) resulting in a complex population index bias.

4.3.3. Optical Distortion

Besides the vignetting and projection cut-off effect, distortions are a second effect due to the intrinsic nature of optical systems. The distortion of an image becomes worse to the edges and corners of an image and affects especially wide-field optics. CILBO has a FOV of 22×28 deg^2, resulting in a total area of 616 deg^2. With a half-sphere area of around $20,625$ deg^2, CILBO covers only 3 % of the night sky. Although the system does not perform an All-Sky surveillance, MetRec still needs to solve the image plate.

For this purpose, MetRec uses a 3$^{\text{rd}}$ order polynomial function. Based on this fit, Albin et al. [2017] performed an analysis on the distortion correction. Not accurately computed astrometric data can lead to a systematic bias in the orbit and radiant determination of meteors.

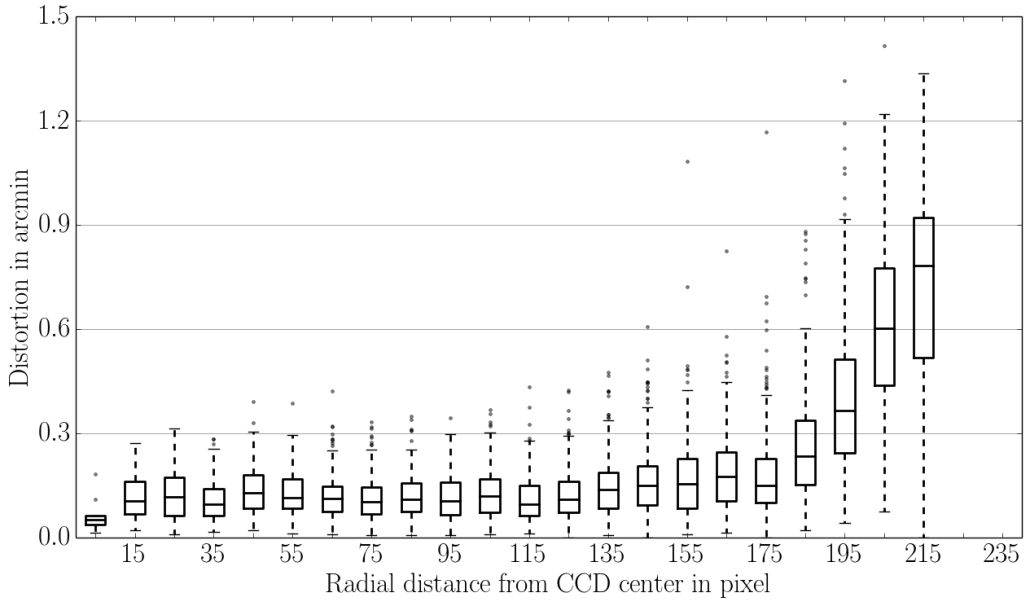


Figure 4.12.: Optical distortion of the meteor camera systems (merged data set) in arcminutes vs. the radial distance from the Charged-Couple Device (CCD) centre respectively optical axis in down-sampled pixels. The data are shown as a box plot. Each box includes data within a range of 10 pixels. From Albin et al. [2017].

To analyse the plate solving performance, several randomly chosen observation night data are used. The actual position of the visible stars (catalogue based) is compared with the determined position from MetRec. With the deviations, a distortion map plot can be computed. This map is a vector field graphic, where the length and direction indicates the distortion direction and magnitude at certain positions on the CCD chip. Based on these maps, figure 4.12 is computed that shows the distortion in arcminutes vs. the radial distance from the centre of the CCD respectively the optical axis. The distance is given in down-sampled pixels and the results are represented in box-plots. Each box represents the data for a bin with a width of

10 pixels. For simplifications, only the values along a virtually set horizontal line (crossing the centre of the CCD) are considered.

The graph shows that the distortion remains quite constant between the optical axis and a radius up to 140 pixels. The median stays at around 0.1′. The down-sampled, virtually set horizontal line on the CCD has a length of 388 pixels. Thus, along a horizontal length of around 80 % of the total chip the distortion is not significant. 95 % of all data have a distortion of around 0.2′. To the edges and corners, the values worsen up to 0.75′ and more.

In conclusion, MetRec's 3rd order polynomial corrects the optical distortion sufficiently up to 80 % away from the optical axis. Position measurements beyond that should be considered carefully, since the resulting bias affects orbit and radiant computations. An automatic correction algorithm for these data is currently neither implemented nor planned.

4.3.4. Astrometric accuracy

Besides the correction of the optical distortion, MetRec determines also the uncertainties of the measurements and stores values in the log-files. This *accuracy* refers to the so-called *goodness-of-fit* and describes the root mean square of all deviations between the photometric centres and a fitted projected trajectory line (however, hereafter it is called accuracy). This line describes the meteor on a great circle, to consider the movement of a sphere properly. Furthermore, MetRec applies also a velocity function, to project proper meteor positions on the great circle.

To analyze MetRec's performance, figure 4.13 shows the normalized (sum of all histograms bins equals 1) distribution of the accuracy values in arcminutes. The yellow and blue bars show the results for ICC7 and ICC9, respectively and the bin width is set to 0.1′. This value is the according minimum accuracy resolution of the detection software. The histogram shows the distributions between 0.0′ and 4.0′. Around 3 % of all data are computed with a worse accuracy, up to 10′, however, for a proper display, the graph is clipped.

Both CILBO distributions consist of two separate parts: A skewed distribution with a maximum at around 0.7′ to 0.8′ and a single peak at 0.0′, indicating a perfect correspondence between the fitted line and the actual measurements. For ICC7 around 8 % and for ICC9 around 11 % of all data lay in this bin. Since MetRec rounds the computed accuracy values, a fraction of the 0.0′ values can be explained by calculation or rounding errors. However, most "perfect" data appear to be wrongly determined, since the lack of data in the bins 0.1′ and 0.2′ can only result from an algorithmic error. Future software updates should consider the accuracy determination.

Not considering the erroneous data at 0.0′ the median and corresponding Inter Quartile

Range (IQR) for ICC7 and ICC9 are:

$$acc_{\mathrm{ICC7}} = 1.2^{+0.9}_{-0.5}{}'$$
$$acc_{\mathrm{ICC9}} = 1.0^{+0.5}_{-0.3}{}'$$

The median values correspond to an accuracy of about a quarter CCD pixel and considering the large FOV and the underlying interline transfer readout these results are feasible for proper scientific analysis. These values are e.g., used for the MOTS Monte-Carlo extension [Albin et al., 2016]. The extension uses the accuracy values as an input for the Monte-Carlo computation, since the astrometric centres of the meteors cannot be determined perfectly. However, the erroneous $0.0'$ data need to be replaced by an average value.

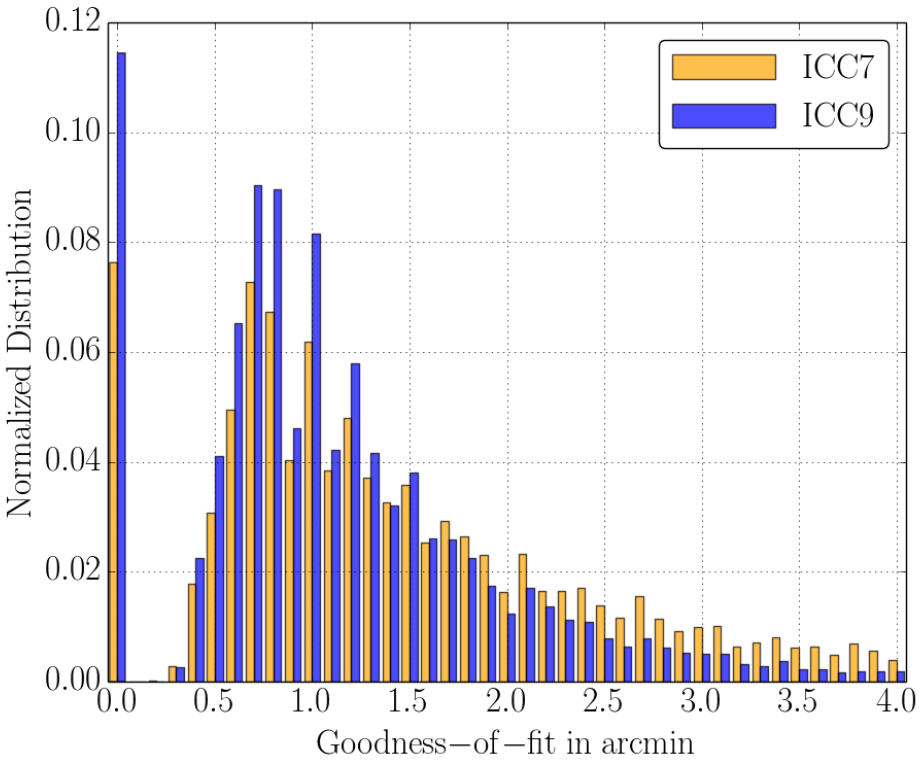


Figure 4.13.: Normalised distribution of the accuracy respectively goodness-of-fit in arcminutes. The yellow and blue bars indicate the data for Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9, respectively. Each bin has a width of $0.1'$. From Albin et al. [2017]

Further insights of resulting bias effects of the MetRec software can be revealed by analysing the accuracy vs. the number of recorded frames per meteor. The minimum number of video frames is 3. Meteor appearances on less frames are not considering / stored. Figure 4.14 shows a box plot, where the accuracy data is plotted vs. the number of recorded video frames. Note: the erroneous $0.0'$ data are not used in this analysis. The longest meteor appearance is 80 frames long, so, this figure shows the majority of the data between 3 and 40 frames for a clearer visualisation and merges the dataset of both camera stations. A number above represents the

number of accuracy values per box.

The median accuracy and corresponding width of the IQR worsen from 3 to 7 frames, but the number of outliers outside the whiskers decrease. In rare cases, the accuracy is worse than 8.0′ and more for 3 video frames long meteors. Between 7 and 8 frames, the median and IQR improve again and worse slightly again to 40 frames.

The improvement between the 7$^{\text{th}}$ and 8$^{\text{th}}$ frame is due to MetRec's velocity fitting algorithm. Up to 7 frames, the software uses a constant velocity fit, assuming the velocity does not change significantly within the few hundred milliseconds of appearance. The large scattering of the accuracy indicates that this assumption does not perform well. 8 frames and more are fitted by a linear velocity function. This appears to be more feasible and the accuracy improves again. Longer meteor records correspond to a long atmospheric entry phase. The meteor decelerates significantly; the drag is proportional to the squared velocity of the meteor. Thus, the linear velocity fit becomes insufficient again for very long meteors and the accuracy drops again, however not as strongly as for a constant fit function.

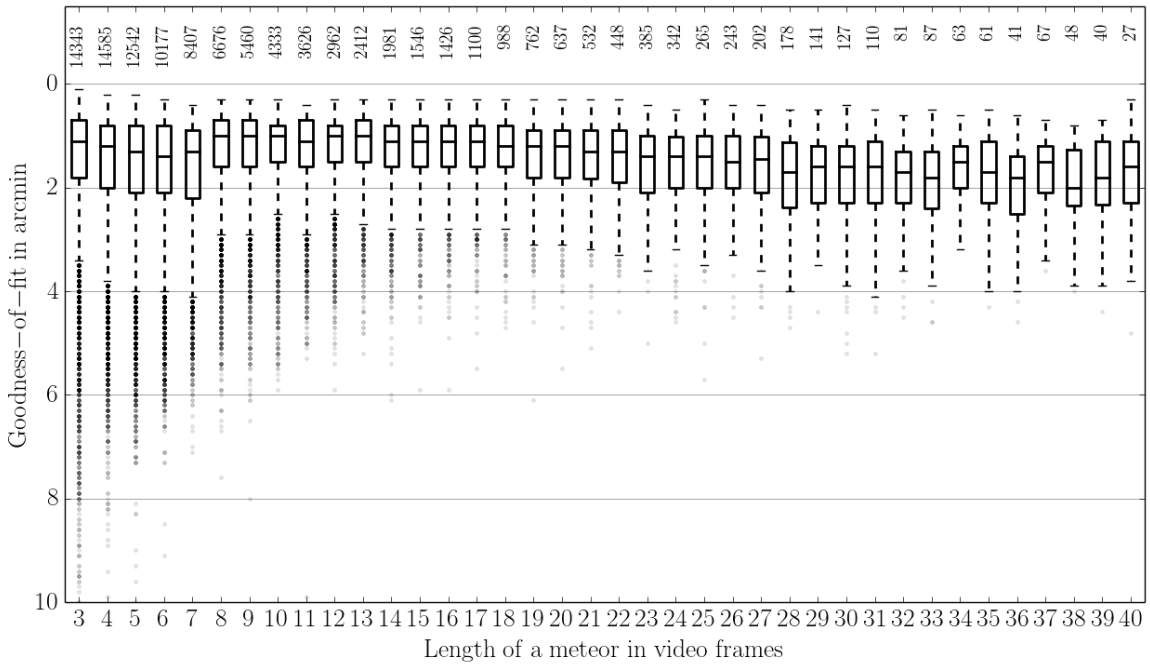


Figure 4.14.: Accuracy / goodness-of-fit in arcminutes vs. the length of a meteor's appearance in number of video frames. The figure shows the data in a box plot. Furthermore, the data from both cameras are merged. From Albin et al. [2017].

In an additional approach, the accuracy is analysed by Albin et al. [2017] considering the brightness of a meteor (for further details on the recorded brightness distribution see section 4.3.7). Figure 4.15 shows a box plot, where the accuracy in arcminutes is plotted vs. the peak brightness of the meteors for ICC7. The plot shows the data from -2.0 mag to 5.5 mag and each data bin has a width of 0.5 mag. The number of data points per bin is listed above each box.

The median and corresponding IQR improve for fainter meteors: -2.0 mag bright meteors

are determined with an accuracy of around $3.0' \pm 1.0'$ and 5.5 mag meteors have an accuracy median of less than 1.0'. Since ICC9 shows a similar distribution, the data are not shown (see Albin et al. [2017]). In average the accuracy medians are better for brighter meteors by around 1 arcminute.

Bright meteors overexpose the CCD pixels. This leads to a "charge-overflow" and thus blooming, leading to a broadening of the meteor's appearance. Consequently MetRec cannot determine the photometric center correctly, causing a severe bias in the accuracy. Another effect may be the bias / shift due to a wake. With a population index r of around 2.2 the rather bright and wake developing Perseids are affected most.

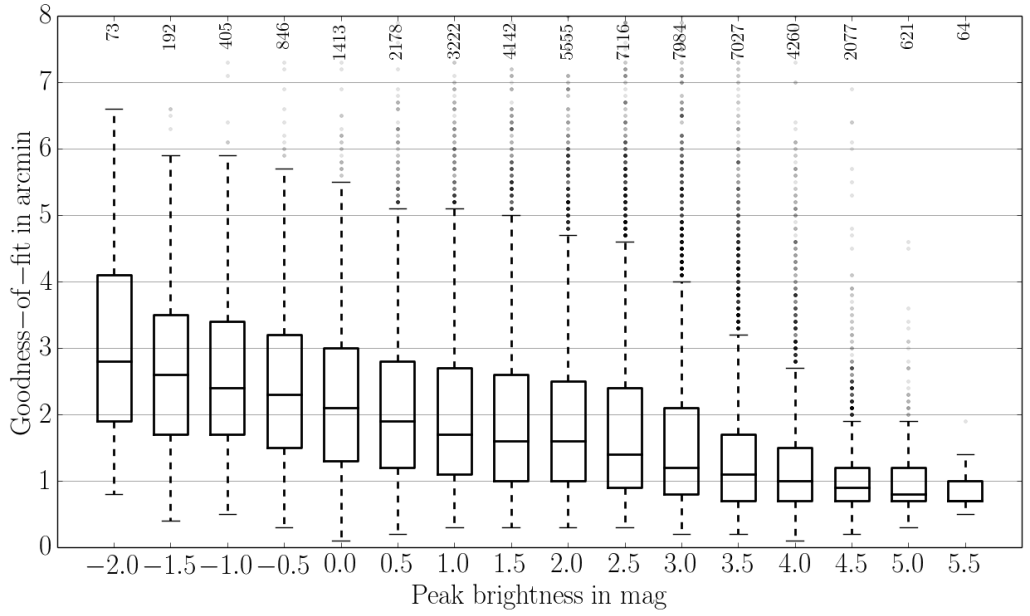


Figure 4.15.: Accuracy / goodness-of-fit in arcminute vs. peak brightness in magnitudes for Intensified CCD Camera (ICC)7. The accuracy ranges from $0.0'$ to $8.0'$ and the data is represented in box plots. Each box includes a data within a range of 0.5 mag, showing the results from -2.0 mag to 5.5 mag. From Albin et al. [2017].

In conclusion, to simplify the Monte-Carlo extension of MOTS, one could use simply an accuracy value between $1.0'$ and $1.2'$ as suggested by Albin et al. [2017].

4.3.5. Video Frame Distribution

As mentioned MetRec considers only meteors that appear on at least 3 consecutive video frames. In figure 4.14 one sees that the accuracy is depending on the number of recorded frames. Further analysis in this section consider the video frame number as an important parameter for correlation analysis of different measurements. The number of recorded video frames corresponds with the so called appearance time of a meteor: Since CILBO is operated with 25 frames per second, each video frame corresponds to an exposure time of 40 ms. Assuming that a meteor appears at the end of the first and disappears at the beginning of the last frame, respectively, a meteor's appearance time on 3 video frames does not correspond consequently

to 120 ms. In fact, the appearance time can be close to 40 ms. This strong variation affects the velocity determination for short frame meteor recordings and is analysed in a later section. This effect is less relevant for longer recordings.

Figure 4.16 shows a histogram of the number of video frames for ICC7 (solid line) and ICC9 (dashed line). The plot ranges between 0 and 20 frames. Meteors up to 80 frames length have been recorded, but for a better readability, the histogram is clipped. Furthermore, in this case only simultaneously detected meteors are used [Albin et al., 2015a], namely 11,785 recordings. The distribution for ICC7 data has its peak at 3 frames and drops for longer meteors. ICC9 on the other hand records most simultaneously detected meteors on 5 frames long videos and drops significantly for 7 frames long recordings. An explanation is provided in a later section, where the measurements are correlated with each other.

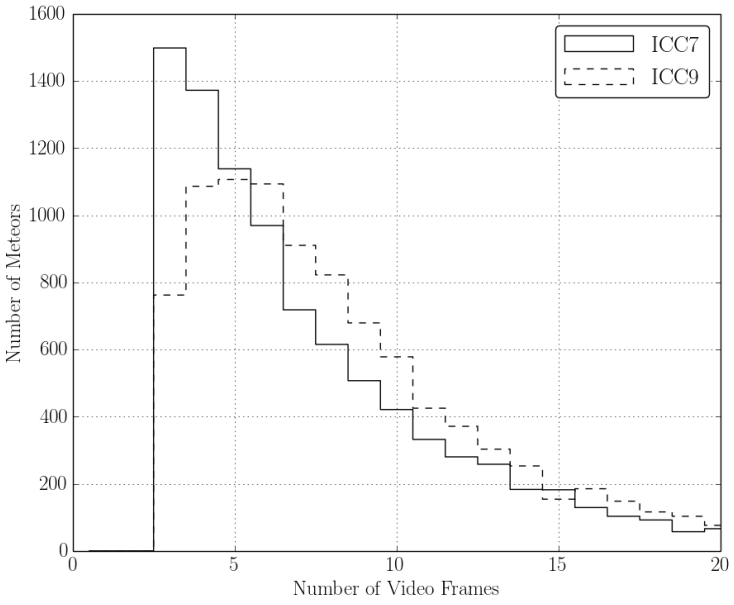


Figure 4.16.: Absolute distribution of video frames number. The solid and dashed lines indicate the data for Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9, respectively. Here, only data from simultaneously observed meteors were used. From Albin et al. [2015a].

4.3.6. Apparent Velocity

Figure 4.11 shows the *signal ratio* of the CILBO system, vs. the apparent velocity in pixels per second. The analysis uses the apparent CCD chip velocity, for a more detailed and sophisticated analysis. Understanding the velocity distribution based on technical bias effects is an important issue for scientific interpretations and implications.

Figure 4.17 shows the absolute distribution density of the velocities from the inf-files. The plot shows the data from 0 pixels per second to 300 pixels per second. Note: with the CILBO set-up, a velocity of 100 pixels per second corresponds to an approximate angular speed of around 7 degrees per second. Yellow and blue lines indicate the data for ICC7 and ICC9.

The solid lines represent all data. Both distributions have a similar shape with a peak at around 35 pixels/s for ICC7 and around 48 pixels/s for ICC9. The distribution representing the ICC9 data converges faster to 0 than the Tenerife set-up. Since the ICC7 distribution is flatter after the peak, it crosses the other function at around 195 pixels/s. Median velocity $\tilde{v}$ and corresponding IQR for both are:

$$\tilde{v}_{\mathrm{ICC7}} = 158^{+151}_{-077} \text{ pixels/s}$$
$$\tilde{v}_{\mathrm{ICC9}} = 146^{+093}_{-066} \text{ pixels/s}$$

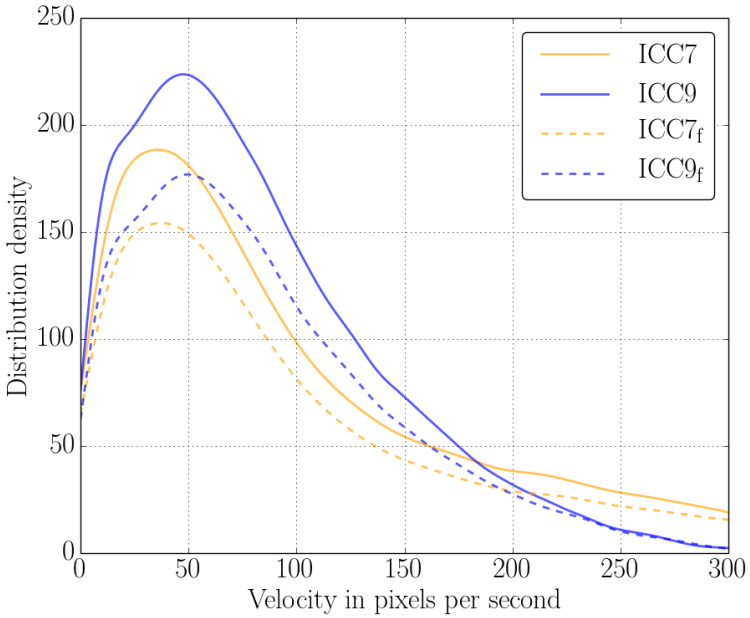


Figure 4.17.: Distribution of the apparent meteor velocity on the Charged-Couple Device (CCD) chip in pixels per second. The distribution is not normalised and indicates an absolute data density. Yellow and blue lines show the distributions for Intensified CCD Camera (ICC)7, respectively Intensified CCD Camera (ICC)9. The solidly drawn distribution shows all data and the dashed ones show a filtered data set, where the first and last determined velocity value is neglected (velocity determined between first and second, and second to last and last video frame. Further explanations in this regard are written in the main text). From Albin et al. [2017].

As already mentioned, one cannot determine the exact appearance and disappearance time of a meteor in the first and last frame. Each frame has a length of 40 ms, where the meteor can (dis-)appear randomly. Consequently, the angular distance between the photometric centres of the first to second, and second to last to the very last frame is underestimated. Filtering these data out leads to the dashed distributions shown in figure 4.17. The peaks and the shapes remain approximately the same. The median and IQR values shift to:

$$\tilde{v}_{\mathrm{ICC7,f}} = 157^{+149}_{-076} \text{ pixels/s}$$
$$\tilde{v}_{\mathrm{ICC9,f}} = 150^{+095}_{-067} \text{ pixels/s}$$

The effect of biased velocity values at the beginning and ending of a meteor recording is affecting the derived physical parameters as described in Albin et al. [2015b]. Several recordings show an increasing velocity, as expected, between the first and second determined speed value (e.g. figure 4b, 6, 7a, 8b and 9c in Albin et al. [2015b]). Thus, the initially entering velocity value cannot be determined on the observational data.

In the following sections, if not mentioned otherwise, the filtered velocity data set is used.

To obtain a more sophisticated view on the velocity distribution shift to slower angular CCD velocities (see figure 4.17), Albin et al. [2015a] performed an analysis considering the number of video frames, similarly as in the accuracy section. Figure 4.18 shows the angular velocity in relative CCD coordinates (see section 4.3.2) vs. the appearance time of a meteor in ms. 40 ms correspond to the length of one frame. Each dot represents the mean result for a certain appearance / detection time. The 1 sigma error bar varies between 0.1 and 0.2 u/s and is not displayed for better readability[4]. Circles and squares indicate the data for ICC7 and ICC9, respectively. A dashed and solid line show the mean GP fit with a grey area as the confidence interval, to interpolate between the discrete appearance times, properly. Due to the lack of data, the fits are not sufficient beyond the given data bins.

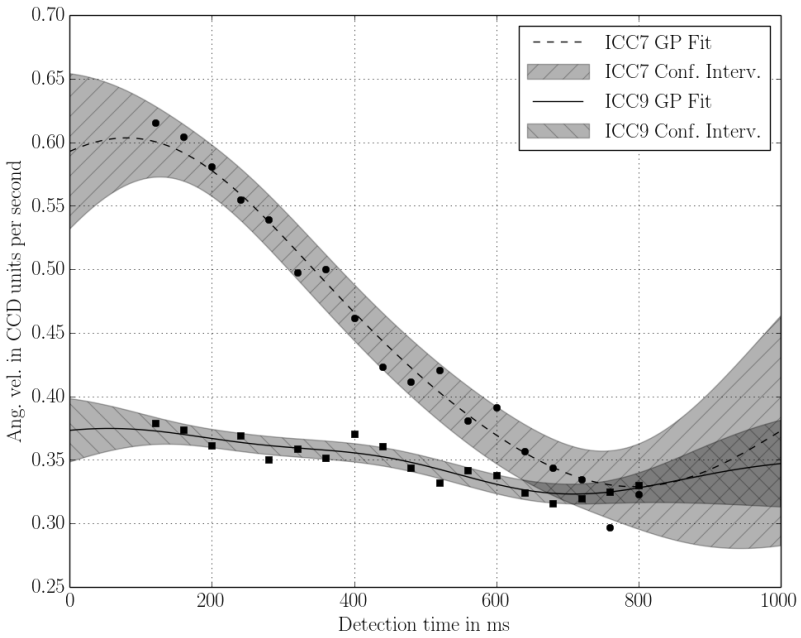


Figure 4.18.: Apparent velocity of the meteors vs. the appearance time in ms. 40 ms correspond to the time window of one video frame. Circles and squares indicate the mean data for Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9, respectively. The 1 sigma error bars vary between 0.1 and 0.2 u/s and are not shown to improve readability of the plot. Solid and dashed lines with a gray area show the mean and confidence interval based on a Gaussian Process (GP) fit. From Albin et al. [2015a].

High speed meteors correspond to short appearances. Longer meteors appear slower on the ICC7 camera (up to a factor of 0.5). The slope for the ICC9 data is less prominent and decreases

[4]Note that 1 u^2 represents the normalised CCD area

in total by approximately 10 %.

Both cameras are built similarly. The slight offset in the optics cannot explain this biased velocity measurements. Reasons for this bias are provided in the time dependency and pointing influence section, respectively.

4.3.7. Brightness

The ICC sub-systems are similarly built, but show deviations in miscellaneous statistics (e.g., different accuracies). Here, the brightness distributions are compared with each other to identify possible deviations and biases, too. Brightness measurements affect the accuracy, respectively the goodness-of-fit as mentioned in section 4.3.4 (figure: 4.15).

Assuming that meteors appear randomly on sky, meteors also appear and / or disappear outside the FOV of CILBO [Albin et al., 2017]. If the maximum brightness occurs outside the FOV, the peak brightness distribution shifts. For the analysis only these meteors are considered that have been completely observed. Therefore, only these meteors are considered that do not cross the 5 % margin at the edges of the CCDs. This leads to a smaller rectangular FOV. Applying this constrain on MetRec's inf-files, around 3 % of the ICC7 and 4 % of the ICC9 data are sorted out.

Figure 4.19 shows a normalised distribution of the peak brightness of the filtered dataset. The orange and blue lines represent the ICC7 and ICC9 distributions. To obtain these continuous functions, a KDE is applied on the data sets to smooth the data. The brightness is shown within a range of 0.0 mag to 6.0 mag. Both distributions deviate from each other. The peaks for ICC7 and ICC9 are at 3.1 mag and 3.5 mag, respectively. ICC7 is shifted towards brighter meteors while ICC9 detects more fainter ones. However both functions converge to null at 6 mag; the approximate detection limit of the system. The apparent peak medians and corresponding IQR ranges are:

$$m_{\mathrm{ICC7,V}} = 2.92^{+0.76}_{-0.97}\ \mathrm{mag}$$
$$m_{\mathrm{ICC9,V}} = 3.32^{+0.70}_{-0.88}\ \mathrm{mag}$$

Despite two similar settings, ICC9 detects in average around 0.40 mag fainter meteors than ICC7.

To determine biases during the measurement of a meteor the filtered brightness data is analysed vs. the appearance time in ms. Figure 4.20 shows the mean peak brightness vs. the appearance time in ms. For the analysis, the data is divided into sub sets with the same number of recorded video frames. The shown appearance time range is shown from -125 ms to +125 ms. 0 ms indicates the centre of the meteor recording (centre frame for uneven number of frames). Four different symbols represent the mean values for ICC7: circle: 3 frames long, triangle: 4 frames long, star: 5 frames long, diamond: 6 frames long. For a better readability the 1 sigma error bars are not displayed. Furthermore, the symbols are connected with dashed lines and do

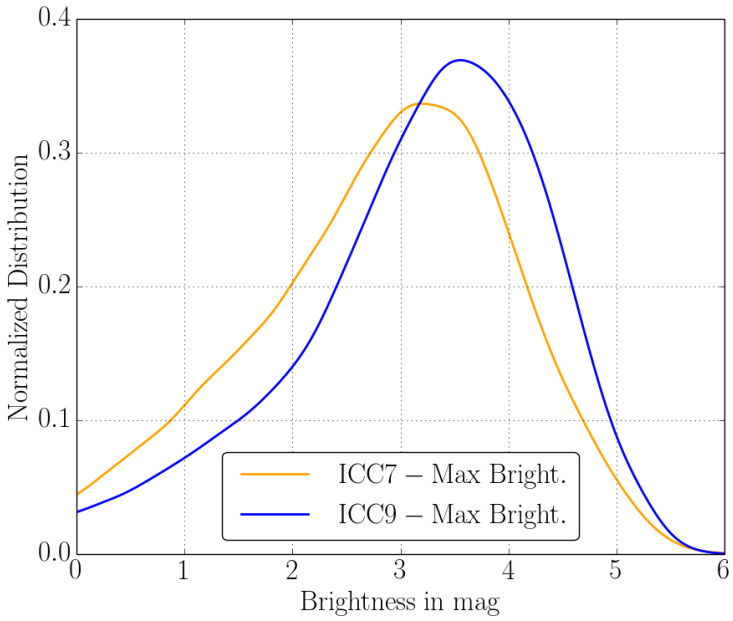


Figure 4.19.: Peak brightness distribution of the recorded meteors. The yellow and blue line show the normalised distribution functions for Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9, respectively. The functions were determined using a Kernel Density Estimator (KDE); and only meteors are considered that are observed completely (5 % margin with respect to the edges of the sensor chip). From Albin et al. [2017].

not represent any physical function or relation between appearance time and brightness.

The shown sub-data sets contain consequently the light curve profiles of sporadics as well as of shower meteors. Merging miscellaneous populations, however, does not lead consequently to an insufficient analysis. The curve progressions of the shown data are approximately symmetrical. The shape of meteor light curves are described by the dimensionless and commonly used F parameter, where t_{B} and t_{E} are the beginning and end times of a meteor trail, where the brightness is one magnitude less than at the time of its maximum brightness $t_{\mathrm{m,max}}$:

$$F = \frac{t_{\mathrm{B}} - t_{\mathrm{m,max}}}{t_{\mathrm{B}} - t_{\mathrm{E}}}$$

Previous works and research on meteor light curve profiles revealed that most shower and sporadic meteors have an average value close to $F = 0.5$; corresponding to a symmetric profile [Brosch et al., 2004, Koten and Borovička, 2001]. Airborne based observations of the Leonids have been done by Murray et al. [1999], who confirmed the symmetry for this particular shower. Table 2 in Koten et al. [2004] summarizes several F parameters and demonstrated that the value of 0.5 applies mostly perfectly. F does not contain any information on the width of the profiles, however here in this work it is determined whether CILBO observes in average also a symmetric profile or whether its biased. An average profile is determined with a GP fit in figure 4.20 that shows that for ICC9 meteors with a frame length of 6 frames appear symmetric.

In average, meteors that are recorded on less video frames, are fainter than longer appearances. This appears reasonable, since fainter meteors reach the detection threshold of MetRec only for a short time. The brightness of larger meteors increases over a longer appearance time (while the recording is triggered), before it faints after its peak. Moreover, the average profiles appear symmetrically, although concluding a symmetry from only 3 frame long appearance is rather undetermined.

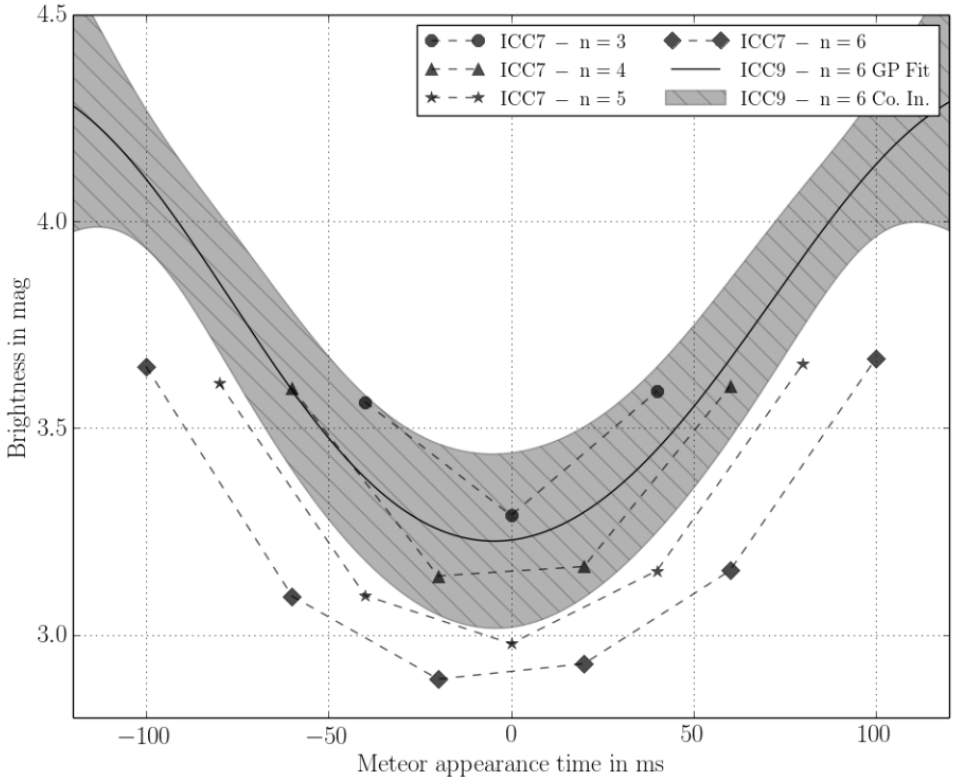


Figure 4.20.: Average brightness in mag. vs. the appearance time in ms. The data set is divided in sub-sets with identical video frame counts. Brightness and appearance time are given in mag and ms, respectively. The data points (circle, triangle, star and diamond shapes are encoded in the legend on the top right) represent the average values for Intensified CCD Camera (ICC)7. A Gaussian Process (GP) mean fit (black solid line) with a corresponding confidence interval (grey hatched area) shows a regression attempt for thee 6 frames length meteors light profiles; recorded with Intensified CCD Camera (ICC)9.

Additionally, the average profile that is corresponding to a length of 6 video frames is compared with the average profile from the ICC9 station to determine any inter-station biases. For the comparison a GP fit is applied on the 6 frames long ICC9 light curve profiles. This demonstrates the feasibility of GP fits for further scientific research in this regard. A black line and hatched grey area indicate the mean fit and corresponding confidence interval of the 6 frames long ICC9. For better readability, the actual data points are not shown. The GP algorithm finds a nicely shaped symmetric function to describe the light curve profile. Although the data gaps are rather large (40 ms) the GP determines due to the symmetry of the data a feasible interpolation. The non-availability of further data points beyond the beginning and ending

frames leads to a broadening and worsening, however the interpolated values up to ±125 ms appear still symmetric and extrapolate the curve progression (the solid mean fit) well. Thus, with a GP fit, the data gaps can be interpolated and the profile can be extended in this case by about 25 %.

Compared with the ICC7 data, the mean fit of the ICC9 measurements indicate 0.5 mag fainter meteors with a frame length of 6. Also, this inter-station bias or shift is explained in the following sections regarding the time dependency and pointing depending bias effects on meteor cameras.

4.3.8. Observation time and pointing effects

Both cameras are similarly build: same camera and optical system, identical electronics, housing and software tools to detect, record and store meteor observations accordingly. ICC7 and ICC9 are mounted on Tenerife and La Palma, aiming approximately towards the east and west, respectively. The boresights have an elevation of 53 degrees with respect to the horizon.

Though, previous sections show biases in the accuracies, recorded frame distribution, brightness and apparent velocity on the respective CCD chips. ICC7 detects brighter and faster meteors than the La Palma station.

To understand these effects, Albin et al. [2015a] and Albin et al. [2017] considered the pointing direction of both cameras. Figure 4.21 and 4.22 show miscellaneous parameters and measures vs. the observation time for ICC7 and ICC9, respectively. The time is the X-axis, given in UTC from 18 UTC to 06 UTC. The right Y-axis corresponds to the black solid distribution and shows the normalized distribution of detected meteors within the observation nights. For both distributions, a KDE is applied to smooth both distributions.

Both cameras start to detect first meteors around 19 UTC. Note that this time is an average value. The start of the recording depends on the Sun elevation below the horizon and is season depending. Afterwards, both curves increase approximately linearly. ICC7 reaches its peak at around 4.30 UTC, ICC9 around 30 minutes later. Both curves show a few wide and less significant peaks between 20 UTC and 23 UTC. Further, the slope of the ICC9 curve is steeper than for the other camera.

The left Y-axis displays angular values. Note, the values decrease from the bottom to top (200 degrees to 0 degrees). The plot shows the angular distance between the boresight (not considering the FOV of CILBO) and two sporadic sources, namely the Apex (blue) and Antihelion (red) source. Dashed lines with a same coloured area indicate the mean angular distance and the corresponding seasonal variation between sources and boresight. A solid vertical line with an according hatched area indicates the average rise or dawn time of the Apex and Antihelion source with the seasonal variations.

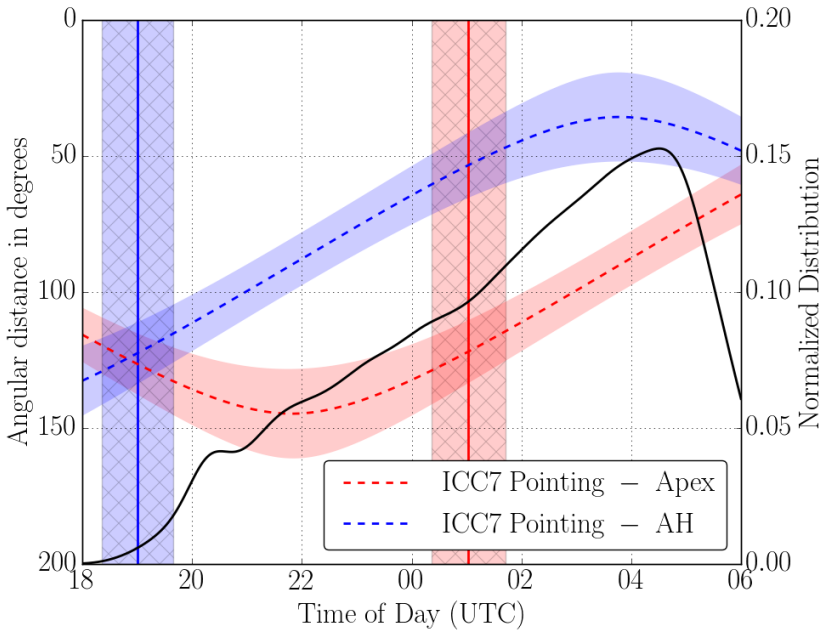


Figure 4.21.: Normalized detection distribution (right Y-axis) and angular distance (left Y-axis) vs. the detection time in UTC (X-axis) for Intensified CCD Camera (ICC)7. The detection distribution is computed using a Kernel Density Estimator (KDE). The blue and red dashed lines are the angular distances between the Antihelion and Apex source with respect to the boresight vector. A solid vertical line indicates the average rise time of the radiants in the source corresponding color. A corresponding coloured area and hatched area show the seasonal variations during the year. From Albin et al. [2017].

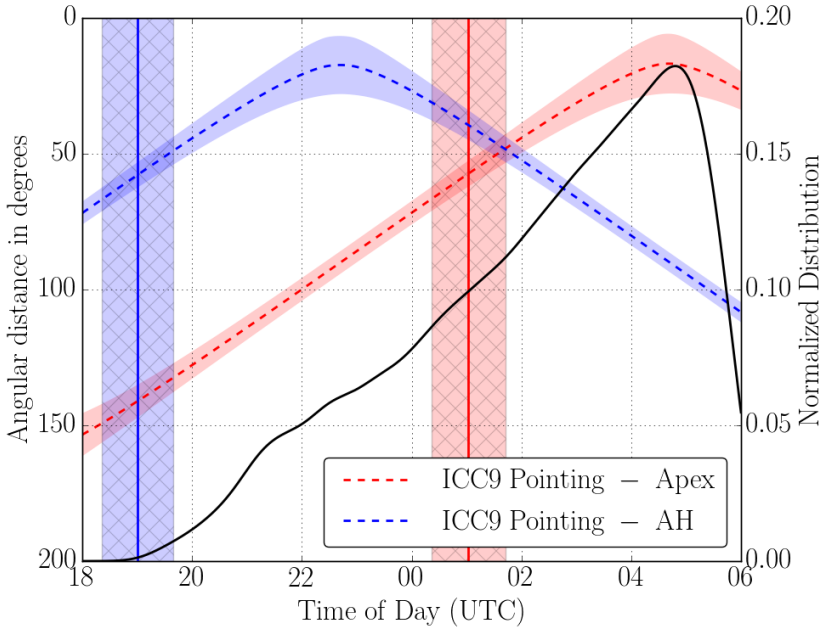


Figure 4.22.: Normalized detection distribution (right Y-axis) and angular distance (left Y-axis) vs. the detection time in UTC (X-axis) for Intensified CCD Camera (ICC)9. The detection distribution is computed using a Kernel Density Estimator (KDE). The blue and red dashed lines are the angular distances between the Antihelion and Apex source with respect to the boresight vector. A solid vertical line indicates the average rise time of the radiants in the source corresponding color. A corresponding coloured area and hatched area show the seasonal variations during the year. From Albin et al. [2017]

For ICC7 the Antihelion radiant rises between 18 UTC and 20 UTC. This corresponds consistently with the set time of the Sun. After the Antihelion dawn, the angular distance between the camera and the source decreases approximately linearly from around 130 degrees to 40 degrees (seasonally down to 25 degrees) between 3.45 UTC and 4 UTC. The Apex source has an angular distance between 110 and 130 degrees at the rise time and decreases also. The minimum is reached at 6 UTC, right before dawn. The peak of the normalized detection distribution is a few minutes ahead of the Antihelion maximum.

The rise times of both sources are slightly shifted for the La Palma camera (figure 4.22), since La Palma is around 1 degree further west. However, ICC7 and ICC9 are aligned to the west and east, respectively. Consequently the angular distances to the sporadic sources appear differently. The Antihelion source reaches its minimum at around 22.30 UTC at around 25 degrees (seasonally down to 10 degrees and less) and increases afterwards to over 100 degrees at 6 UTC. The Apex radiant rises with a small angular distance of 55 degrees and decreases down to 20 degrees before 5 UTC. The minimum corresponds perfectly with the detection maximum.

Both figures show that the number of detected meteors increases over the night and anti-correlates with both sporadic sources. It appears that the Apex source contributes more to the detections than the Antihelion source. Both normalised detection distributions show an increasing slope after the rise of the Apex source. This feature is more prominent for the ICC9 distribution.

To determine the implications of possible biases due to the different pointing directions, first, the apparent velocity and brightness distribution is shown in a time depending figure. For this purpose, the filtered data sets are used. Figure 4.23 shows a density plot, where the velocity in pixels per second is plotted vs. the time. The shown period of time is again 18 UTC to 6 UTC and the velocity ranges from 0 pixel/s to 300 pixel/s, similar as shown in figure 4.17. The figure shows the ratio of the resulting ICC9 and ICC7 distributions that are computed using a KDE. Blue and red colours correspond to a larger percentage of data points in the ICC7 and ICC9 distribution, respectively. White indicates no differences.

The ICC9 velocity bias is present towards slower meteors, especially between 1 UTC and 4 UTC. A second, less prominent feature appears for slow meteors between 4 UTC and 6 UTC for meteors with a speed between 0 and 5 pixels per second. The velocity detection bias can be explained with the different angular distances of the cameras to the sporadic sources. ICC9's boresight is closer to the Apex source than for ICC7. Consequently, meteors that emerge from this region appear slower on the La Palma camera, than for the Tenerife one. A slight increase of the velocity distribution can be explained by detected meteors from the Antihelion source, whose distance increases during the morning hours.

ICC9 detects more slower meteors than ICC7. Slower ones have a shorter angular extent on the CCD chip. Further, the pixel illuminates a pixel longer; the so called pixel dwell time increases. A longer exposure time leads to more induced charges and consequently a better

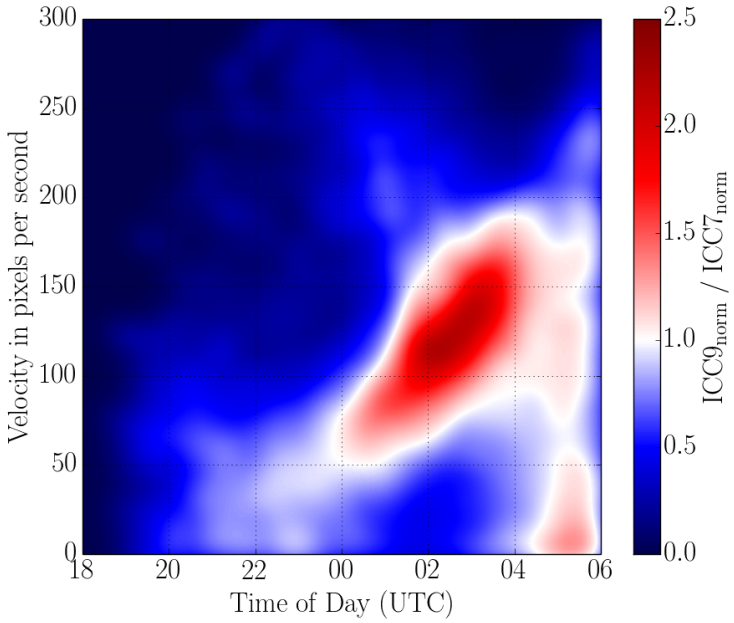


Figure 4.23.: Ratio plot of the Intensified CCD Camera (ICC)9 / Intensified CCD Camera (ICC)7 velocity data in pixels per second (filtered data set), vs. the detection time in UTC. The ratio is colour coded and a bar on the right indicates the ratio. Bluer and redder colours correspond to a higher percentage in the Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9 distribution, respectively. White represents no differences at all. From Albin et al. [2017].

SNR than shorter appearing meteors. Thus, the detectability or detection threshold of meteor cameras is velocity and brightness depending. And since most meteors emerge from the sporadic sources, it is depending on the angular distance between radiant and camera boresight.

The brightness effect can be seen in figure 4.24, where a similarly generated plot is shown as figure 4.23. Here, the ratio of the peak brightness of meteors is displayed. As already described, ICC9 detects more fainter meteors than ICC7 (see figure 4.19). The peak brightness is shown on the Y-axis and ranges from 2.0 mag to 6.0 mag. As expected, ICC9 detects, in comparison with ICC7, more fainter meteors. This detection feature is prominent in the morning hours and correlates well with the decreasing angular distance to the Apex. ICC7 on the other hand detects more faster and brighter meteors. A higher brightness is required to generate a proper SNR for faster meteors and to trigger the detection threshold.

The correlation between the velocity and brightness of a meteor and its resulting SNR explains also figure 4.17 and 4.20. Figure 4.17 shows that ICC7 detects in average faster meteors on short video frames and slower ones on more frames. Longer appearances times correspond to a brighter meteor (figure 4.20) and better SNR. Due to the smaller angular distance between boresight and radiants, the average velocity remains constant for all ICC9 recordings. Figure 4.20 shows further that for the same number of frames, ICC9 detects in average fainter meteors. These faint and slow meteors from the Apex are too fast for the ICC7 system (shorter pixel dwell time), to successfully identify and record these entering particles.

Consequently, ICC9 detects more meteors on less video frames than ICC7. This feature is

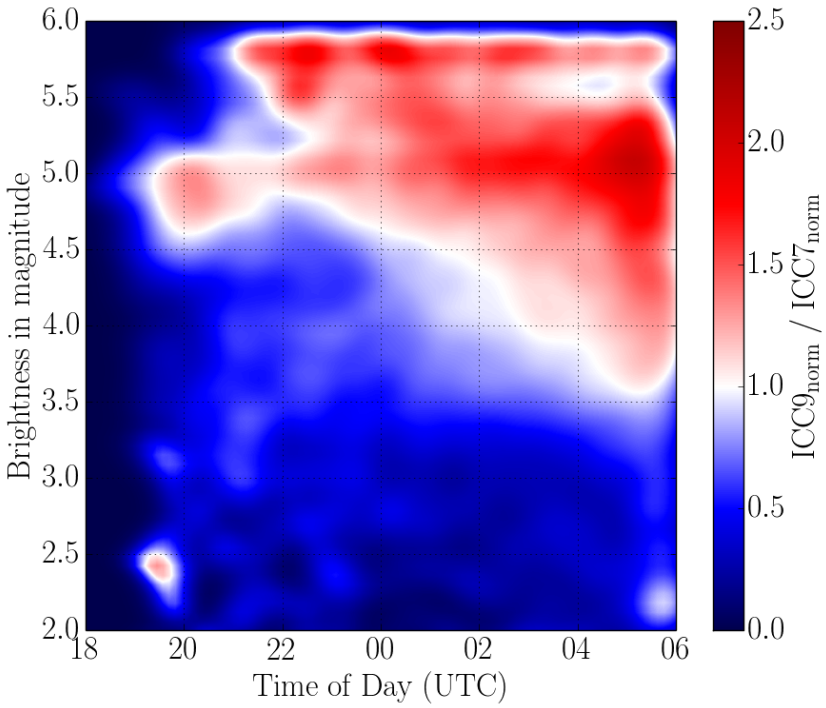


Figure 4.24.: Ratio plot of the Intensified CCD Camera (ICC)9 / Intensified CCD Camera (ICC)7 brightness data in magnitude (filtered data set), vs. the detection time in UTC. The ratio is colour coded and a bar on the right indicates the ratio. Bluer and redder colours correspond to a higher percentage in the Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9 distribution, respectively. White represents to differences at all. From Albin et al. [2017].

very prominent for simultaneously detected meteors and is shown in figure 4.16. Displaying the video frame data separately vs. time leads to figure 4.25. The plot shows three columns with sub-figures. The left and centre image show normalised distributions for each video frame sub-set vs. the detection time in UTC (ICC7 and ICC9). The distribution is color coded and a bar on the bottom indicates the corresponding values. The figure shows the distributions 3 to 40 frames long video frames, however, the number of data leads to suitable distributions up to 30 frames. Furthermore, a third plot on the right shows the ratio of the ICC9 distributions vs. the ICC7 distribution. Each distribution is computed using a KDE.

ICC9 detects more short frame meteors in the morning hours than its Tenerife counterpart. The closer boresight configuration of ICC9 allows to detect faint, slow meteors that are only recorded on a few frames. Longer recordings in the morning are dominated by ICC7. These are the brighter, faster meteors that have a large angular speed and emerge from the Apex. One can see that longer meteors appear in the evening hours of the observation nights. The slightly shifting frame peak (longer videos in the evening, shorter in the morning) correlates nicely with the changing angular distance to the sporadic sources.

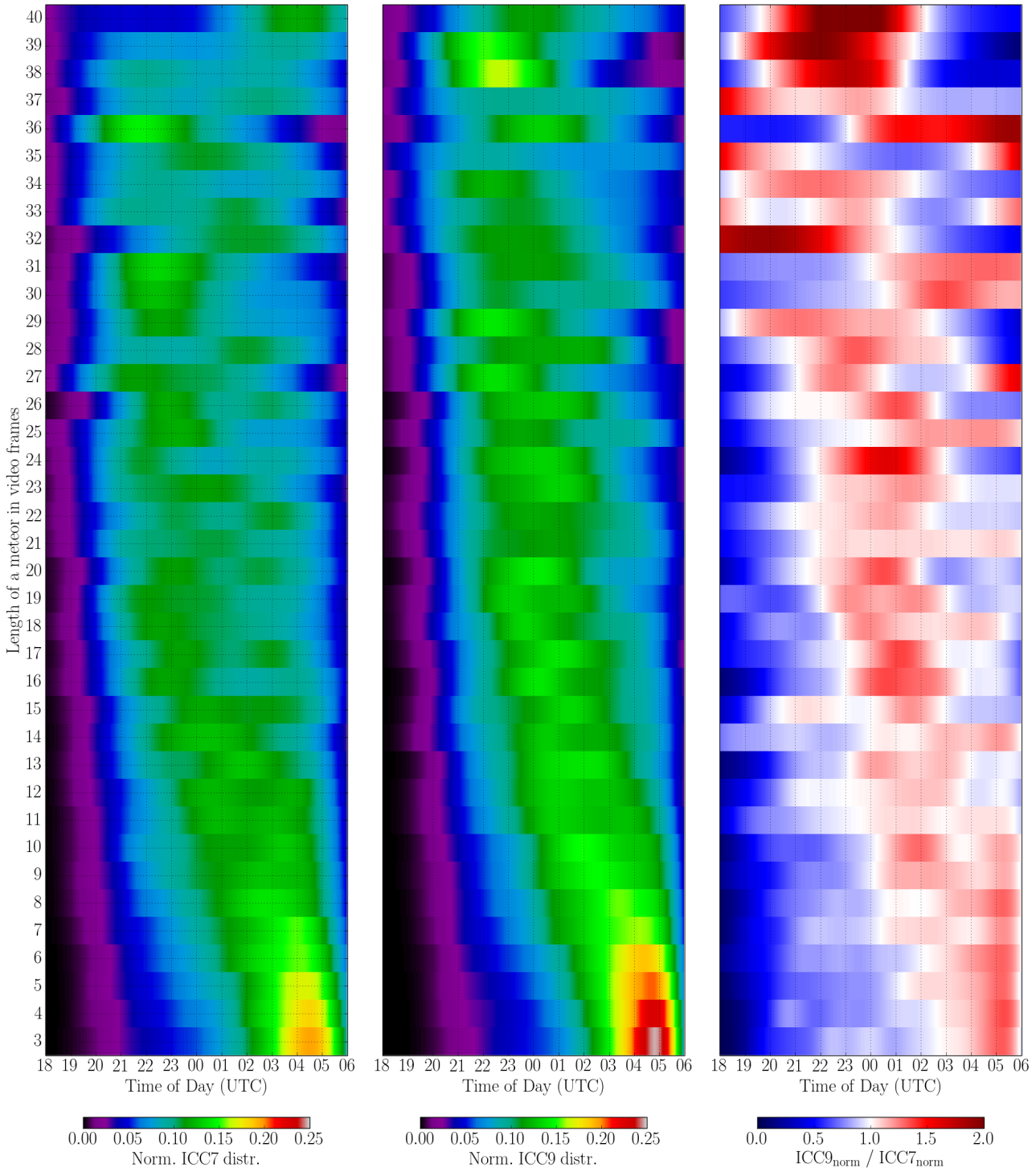


Figure 4.25.: Normalised distribution of the length of meteors vs. detection time in UTC. Each video frame length is shown separately. The left and centre figure show the distributions for Intensified CCD Camera (ICC)7 and Intensified CCD Camera (ICC)9, respectively. The right plot is the corresponding Intensified CCD Camera (ICC)9 / Intensified CCD Camera (ICC)7 ratio plot. Color bars at the bottom indicate the value of the distributions. From Albin et al. [2017].

The time depending variation of the brightness leads also to a time depending accuracy / goodness-of-fit variation (further details: see figure 4.15). Figure 4.26 shows the accuracy in arcminutes vs. the detection time in UTC for ICC7. The data are displayed as box plots, where each box represents the statistics for 15 minutes. The decreasing angular distance between

boresight and sporadic sources, and consequently fainter meteor detections over the night, lead to improving accuracies over the night. The median at 18 UTC is at around 1.9′ and goes down to 1.0′ at 4.15 UTC. Also the total data scattering and width of the IQRs improve accordingly. Since the geometric configuration of ICC9 is from the beginning on better than for ICC7, the improvement is less prominent and is not shown here (see figure 21 in Albin et al. [2017]).

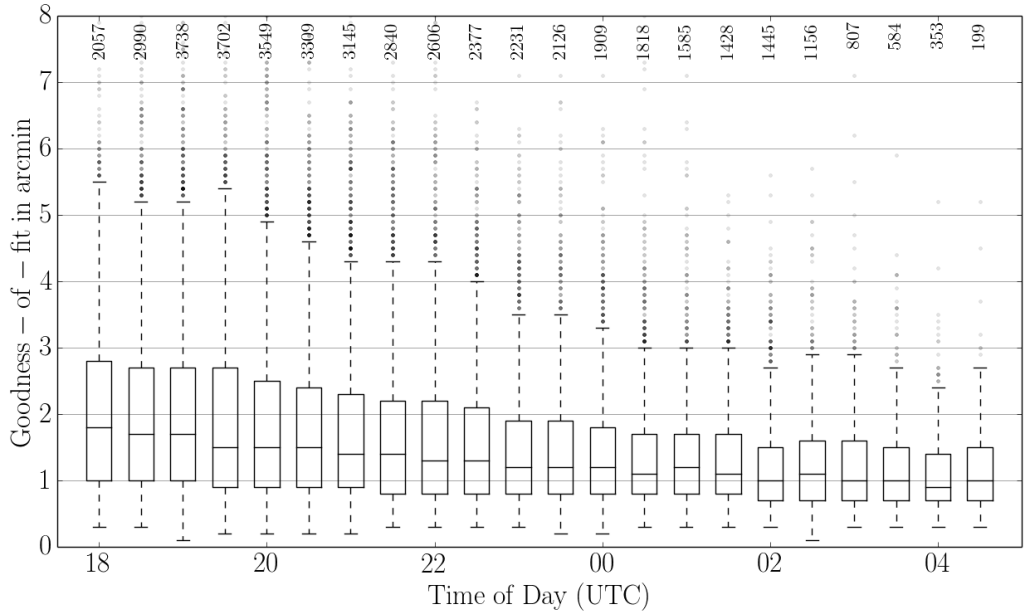


Figure 4.26.: Accuracy / goodness-of-fit of the Intensified CCD Camera (ICC)7 data set in arcminutes vs. the detection time given in Coordinated Universal Time (UTC). The data are shown as box plots and each box considers the data for a period of time of 15 minutes. From Albin et al. [2017].

4.3.9. Conclusion

This section investigated miscellaneous bias effects of CILBO and compared both identically built and operated camera systems ICC7 and ICC9, set up on Tenerife and La Palma, respectively. Previous work were published by the author of this work in Albin et al. [2015b], Albin et al. [2015a] and Albin et al. [2017].

In Subsection 4.3.1 three lists have been provided, describing several expected biases that also apply on other (stereoscopic) meteor cameras: instrumental, algorithmic and statistical biases.

First, the optical system was described and analysed in detail. Subsection 4.3.2 showed that both cameras do not have an ideally aligned optical system. Further, the large FOV and used image intensifier cause a vignetting and illumination drop-off to the edges and corners, respectively affecting the observation of meteors (figure 4.9). Consequently, the deviation between a homogeneously, perfectly observing system and the actual detection distribution was analysed. A new measure was introduced and defined, the so called *detection efficiency* that described this difference. It was found that the parameter depends on the brightness and apparent velocity of a meteor. Faster meteors have a shorter pixel dwell time, illuminate pixels in a shorter period

of time and consequently the corresponding SNR is worse. This effect leads to a technical and scientific bias: the performance of the camera system is underestimated and computed fluxes, mass and velocity distributions are skewed.

Additionally, in subsection 4.3.3 the distortion was analysed. Deviations in the astrometry lead to a worse and wrongly determined trajectories. This effect is very prominent to the edges and corners of the CCD chip, where the distortion becomes around 0.9 arcminutes, compared to 0.15 to 0.20 arcminutes close to the optical axis.

The astrometric accuracy depends on the number of recorded video frames and brightness of a meteor (see subsection 4.3.4). Since MetRec changes its constant to a linear velocity fit, the accuracy improves if more than 8 video frames recorded a meteor successfully. Also, very bright meteors overexpose pixels that leads to an insufficient determination of the photometric centre. Whether distinct wakes or trails cause further accuracy errors, needs to be investigated. It is likely however that wakes cause a shift of the photometric centre that lags behind the actual position of the meteor. Further investigations with more data from the Perseids shower could reveal this effect more clearly.

Further biases in the video frame, velocity and brightness distributions were shown in subsections 4.3.5 to 4.3.7. ICC9 detects more longer meteor appearances; the meteors appear in average slower and fainter compared with ICC7. Considering time dependency it was revealed that these effects are caused by the pointing of the cameras (subsection 4.3.8). Changing angular distances to the Apex and Antihelion sources with respect to the boresights of both cameras cause different geometric observation conditions. ICC9 points closer to the Apex source than ICC7. Consequently, the angular speed on the CCD chip is lower, the pixel dwell time increases and the SNR improves. Even fainter meteors become detectable.

To obtain a proper data set for technical or scientific analysis, the provided raw data need to be filtered. A meteor appears and disappears randomly in the first and last frame of a recording. Thus, the derived velocity value between first to second, and second to last to last video frames are underestimated [Albin et al., 2015b]. So, no proper state velocity can be computed for 3 frames long meteors. 4 frames long meteors provide only 1 suitable value. Longer appearances become more feasible in the determination of the speed. This is considered in section 4.4, where the Monte-Carlo based orbit computation is shown.

Especially the velocity and brightness biases need to be considered for a proper scientific analysis. Neglecting e.g., the brightness bias would shift the derived mass distributions to larger masses and would also affect the population index r. The *detection efficiency* is linked with these effects and additionally biases the flux computations.

Gathering more data will improve and will lead to further analysis. For example, to determine possible temperature effects (weekly, seasonal, yearly) more data are required. Temperature fluxes could affect the accuracy of the camera system and alter the astrometric results.

In general, one cannot apply a simple unbiasing on the scientific data sets. The effects are multi-dimensional and a theoretical dust model should be considered to derive general

rules or a sophisticated, camera properties depending algorithm to unbias the data. For now, certain scientific analysis need to consider the shown effects and apply a statistical correction accordingly.

4.4. Orbit Determination

As already described in section 4.2.4, Meteor Research Group developed a trajectory computation tool for stereoscopic meteor camera systems called MOTS. A detailed description of the software can be found in Koschny and Diaz del Rio [2002] and according manuals. In this section, a Monte-Carlo based extension is described that has been developed by Albin et al. [2016]. Originally, MOTS does only determine a geocentric state vector of a stereoscopically observed meteor. No orbital elements are computed with not corresponding error. This lead to the orbit computation procedure shown here.

The section describes the algorithmic procedure, provides a general overview, how to properly compute orbital elements and gives details in miscellaneous sub-routines that might be of interest for further developments. Results from the technical bias analysis are used to provide new and improved methods for more precise computations.

In an additional section, the resulting database is described. Its content is listed and a few scientific and bias results are shown, too.

Note, the software packages of the Meteor Research Group are currently (end of 2017) developed further by the Meteor Research Group members. The complete Monte-Carlo extension, and MOTS itself are written in PYTHON2.7.

4.4.1. The MOTS Monte-Carlo extension

This section describes in detail the algorithmic approach of the new Monte-Carlo extension software.

Computational / Algorithmic Procedure

The following paragraphs show how the Monte-Carlo (MC) code works. Each paragraph represents a computational step from the detection of a meteor, to the back propagation of the meteor to the Earth's Sphere Of Influence (SOI). The shown procedure is a detailed description of the methods described on page 22 ff in Albin et al. [2016], who summarises the procedure in a processing paper.

1. Identification of simultaneously detected meteors First, a routine compares the data directories of both meteor cameras. The inf-files with corresponding detection times are used to identify possible simultaneously detected meteors. To group the meteor data properly, time stamp and offset need to be corrected, if needed. The studies on the CILBO data has shown that

around 30 % of the complete data set contains stereoscopic observations. The corresponding inf-files are then used for further orbit and dynamics computations.

2. Astrometric data analysis and correction A Monte-Carlo computation calculates a statistical output from a (non-linear) computational problem, considering statistical uncertainties or measurement errors in the computation chain. Here, the astrometric accuracy is the erratic factor in the calculations. Thus, the observation corresponding accuracy value is extracted from the log-files. ICC7 as well as ICC9 store separate log-files. Thus, a stereoscopic observation of a meteor is done with varying qualities. These accuracies are later used to seed the Monte-Carlo algorithm.

Furthermore, the photometric centres need to be unbiased, too. If a meteor crosses close to a similar bright star, MetRec might add the coordinates of the star, too. The software might also wrongly replace the coordinates of a meteor. To remove photometric centres that deviate too much from the possible projected trajectory line, a Random Sample Consensus (RANSAC) algorithm is applied on the data points. Both, the number of photometric centres before and after applying this method are stored in the database. In the next subsection 4.4.2, a short summary provides an overview of the effects of this algorithm. The cleaned data set, containing the astrometric data from the inf-files and the accuracy values from the log-files are then used for further computations. Further, the accuracy value is worsen by 0.3 arcminutes, to consider in a simplified way further effects like distortions etc.

3. Monte-Carlo Seed With the data, the MOTS extension code computes a set of astrometric centres for each camera. This is illustrated schematically in figure 4.27. The first image shows the provided astrometric centers and a possible fit through these data. These means are replaced by Gaussian Probability Density Function (PDF)s, where the standard deviation corresponds to the accuracy value. To generate correct Gaussians, one needs to consider the spherical coordinates of the astrometric data. The angular distance between the right ascension lines converge to the poles. Thus, the width in right ascension depends on the declination.

Based on the Gaussians, a random re-sampling is performed. One possible example is shown in the third image, where red dots indicate re-samples. These re-sampled data correspond to a different fit line (red dashed arrow). For each meteor the re-sampling is done several times, to obtain a set of possible solutions, as shown in the bottom figure (example of 7 possible fits).

4. Rectangular trajectory determination Based on the Monte-Carlo seed, the MOTS trajectory routines are used to determine the trajectory in rectangular state vectors as shown in figure 4.4. Further, the support plane and intersection points are computed w.r.t. ICC7 and ICC9, respectively. Thus, the program determines two trajectories that can be directly compared to verify whether the meteor has been observed stereoscopically or not.

In the section regarding the technical bias effects, a velocity determination bias has been

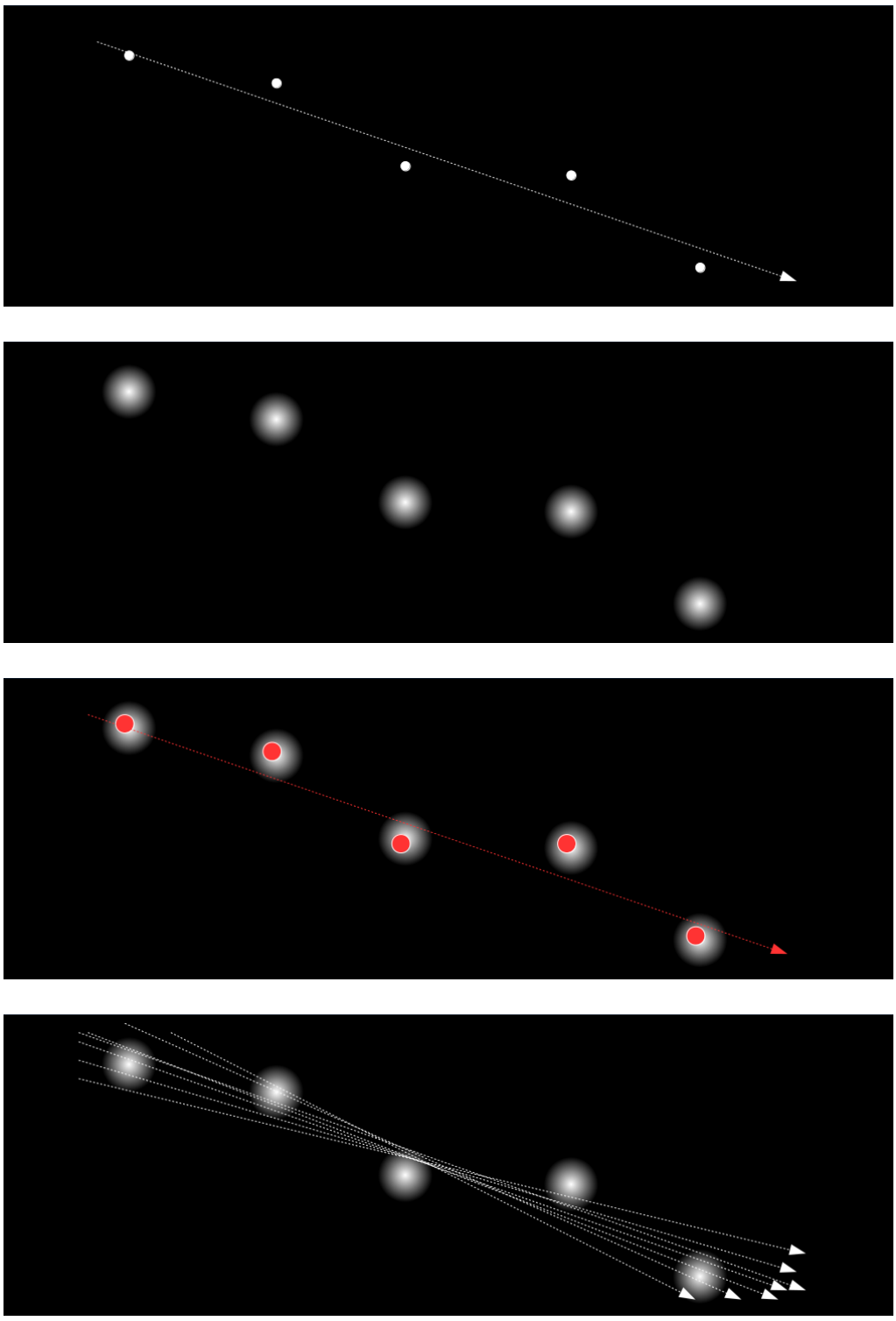


Figure 4.27.: Schematic illustration of the re-sampling / Monte-Carlo seeding. First image: photometric centres of a meteor recorded on 5 video frames with an average fit. The accuracy is not shown, but it is stored in the corresponding log files. A geocentric (here: projected) trajectory shows an average solution. Second image: Replacement of the centres with a Gaussian Probability Density Function (PDF) for each data point. The standard deviation is based on the accuracy value and applies for all Gaussians. Third image: Illustrated re-sampling / Monte-Carlo seed. The red dot represent a re-sampled data set and the arrow is the corresponding the fit using the Meteor Orbit and Trajectory Software (MOTS) routine. Fourth image: A possible set of 7 fit lines, based on 7 re-samples. From Albin et al. [2016].

identified and described that is depending on the number of video frames. Further, due to the random appearance and disappearance of a meteor in the first and last image, the first and last computed velocity values cannot be properly used. Thus, the Monte-Carlo extension software provides the following solutions for a sufficient velocity calculation:

3 Video Frames: No suitable velocity computation. The photometric data are only used to determine the support plane

4 Video Frames: Only one suitable velocity value, derived between frame 2 and 3

>4 Video Frames: Miscellaneous velocity determination methods are programmed and can be applied: Initial velocity (between frame 2 and 3), mean or median of all velocities, or a linear / quadratic fit. In this case, a fit is applied on the data and the computed initial velocity and the beginning of the meteor's appearance is used for the state vector

5. Earth-Centric Coordinate System (ECCS) computation The state vector is then transformed in ECCS. A coordinate system, where the X-Axis point to the vernal equinox and the Z-Axis corresponds to Earth's north pole direction (Y-Axis determined by right-hand-rule). Here, only the Earth is considered as the gravitational centre. Both state vectors are then used to compute the orbital elements. This is done with the Spacecraft Planet Instrument C-matrix Events (SPICE) toolkit[5] and an according PYTHON wrapper[6].

If the entry speed of a meteor does not exceed 11.2 km/s, the resulting eccentricity in ECCS is between 0 and 1. This corresponds to an Earth bound orbit and could be explained by:

- Velocity determination bias. The velocity was not computed properly and lead to an underestimated value.
- Deceleration. The meteor entered Earth's atmosphere already outside the FOV or was not visible at all, since it was too faint. Thus, the computed velocity is not the original orbital speed of the meteor.
- Space debris. The entering particle was a space debris particle that descended.
- "Mini Moon". The meteor was a cosmic dust particle that was captured by the moon; and feel directly on Earth or lost its kinetic energy due to frictions in the exosphere of Earth.

6. Back propagation An eccentricity equal or larger than 1 in ECCS corresponds to an Earth-unbound orbit. In that case, the trajectory is back propagated, until it reaches Earth's SOI at around 900,000 km (see figure 4.28). The computation is adaptive. For each SOI

[5]http://naif.jpl.nasa.gov/naif/index.html
[6]https://github.com/AndrewAnnex/SpiceyPy

crossing the time steps are inverted (negative to positive or vis-versa for positions inside or outside the SOI, respectively) and decreased until the time step is 1 s. This corresponds to an SOI approach of 30 km for a 30 km/s fast particle. At the SOI the state vector, given in ECCS is converted to ECLIPJ2000 and the according orbital elements and further dynamical parameters are calculated.

The SOI of the Moon has a diameter of around 130,000 km. As seen from Earth, this corresponds to an area of around 290 deg^2, respectively 7 % of the whole sky. However, Polle [2017] found gravitational effects and disturbances by the Moon only for close encounters. To simplify the integration and improve the computational performance, the extension package does not consider any moon influences. Future work shall determine the actual number of close encounters.

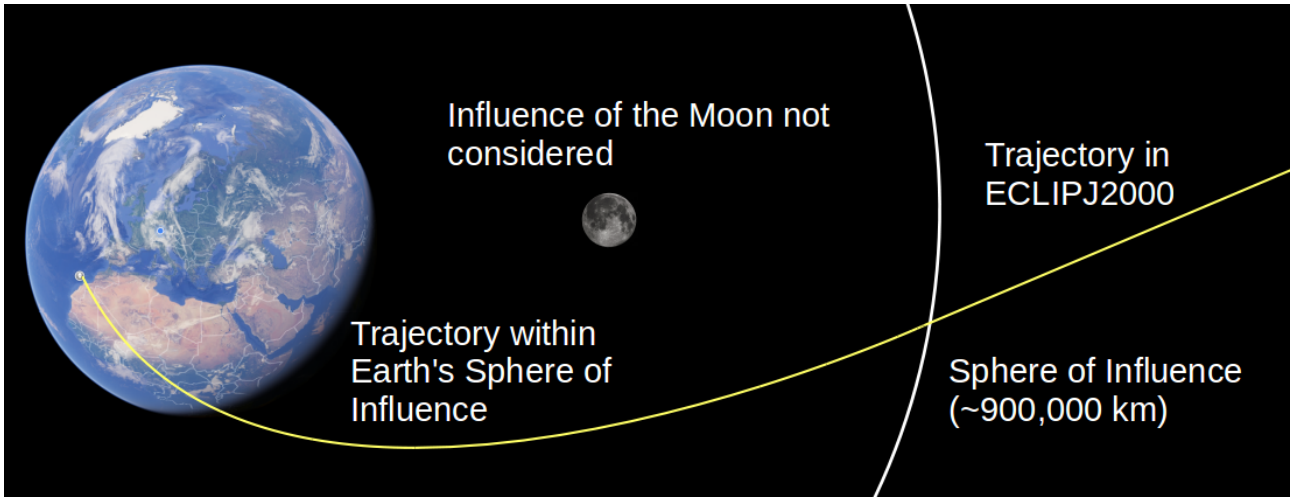


Figure 4.28.: Simplified sketch of the back propagation algorithm. A meteor state vector in Earth-Centric Coordinate System (ECCS) (yellow line) is computed back in time until it reaches Earth's SOI at around 900,000 km. The smallest time step is 1 s. At the Sphere Of Influence (SOI) the trajectory is converted in ECLIPJ2000 coordinates. For the computation the Moon's gravitational influence is not considered. From Albin et al. [2016].

Statistical Consideration

After a complete orbit computation run, the Monte-Carlo method provides a set of possible output parameters in separate PDFs. To store the data in a database properly, some descriptive parameters need to be calculated that describe these distributions. Although most distributions are skewed and cannot be described by a Gaussian or other symmetric functions, internally (in the Meteor Research Group) it has been decided to store unbiased parameters like the median and Median Absolute Deviation (MAD) as well as the mean and standard deviation.

Besides the parameters, an additional probability value needs to be computed. Consider an entry velocity of 11.2 km/s with an error of e.g., 3 km/s. Consequently, a subset of the data describes an Earth-bound orbit, the other subset describes an interplanetary particle crossing Earth's SOI. Thus, each astrodynamic Monte-Carlo method that considers miscellaneous refer-

ence frames and centre bodies, needs to determine probabilities for a certain solution space. In this case, a parameter called *earth_ bound_ orbit_ prob* provides the probability of a meteoroid being on an Earth-bound orbit.

Figure 4.29 shows an output distribution of a recorded meteor at: 2014-08-01T22:54:26 UTC. This test-case is performed with a re-sampling rate of 1,000 and shows the argument of periapsis in degrees in a histogram. Each bin has a width of 3.75 degrees. The solid and dashed lines show the Median and corresponding IQR limits. Median and MAD are 36.1 ± 4.9 degrees and the mean with standard deviation are 37.4 ± 7.5 degrees. The IQR width is 10.1 degrees, with limits at 31.7 degrees and 41.8 degrees. The distribution is slightly skewed to smaller angles and thus not symmetric. The mean is shifted to higher angles w.r.t to the median, since larger values affect the mean more than the Median. The deviation is around 1.3 degrees.

This shows, how the output distribution of the Monte-Carlo computation alters the statistical measures and causes a significant bias between median and mean. For a proper scientific research, the median and corresponding statistical parameters are necessary.

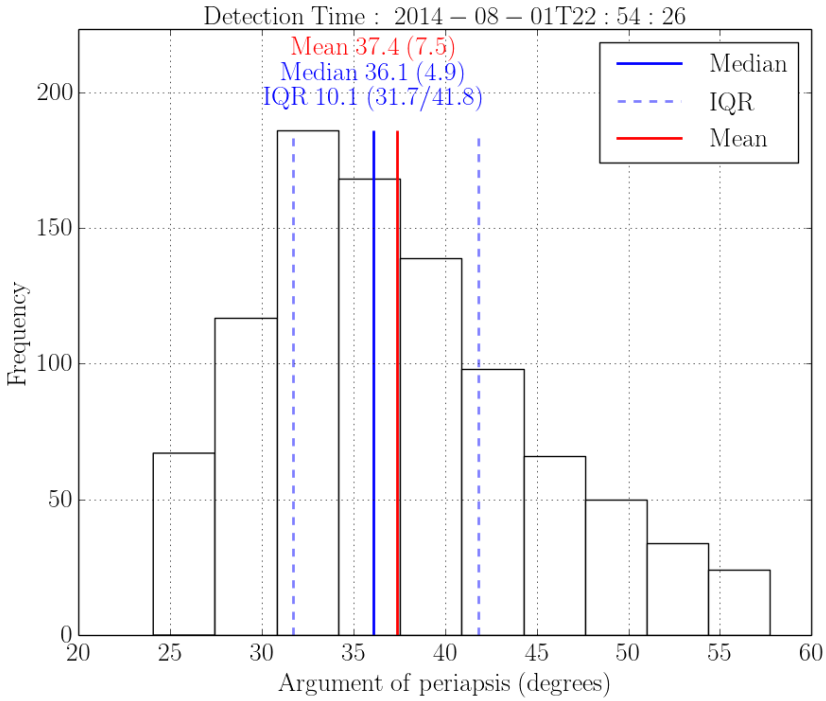


Figure 4.29.: Example of a Monte-Carlo output distribution. The figure shows the results of the argument of periapsis of a meteor at: 2014-08-01T22:54:26 UTC. The sampling rate is set to 1,000 and the frequency of the results is shown in a histogram with a bin width of 3.75°. The solid and dashed blue lines indicate the median and upper / lower Inter Quartile Range (IQR) boundaries. The solid red line shows the mean. On top of the distribution are displayed (from top to bottom): Mean with standard deviation, Median with Median Absolute Deviation (MAD) and width of the Inter Quartile Range (IQR) with corresponding boundaries. From Albin et al. [2016].

Database Set-Up and Cleaning

For each meteor an individual file is stored that contains the statistical results of the Monte-Carlo computation. Each meteor contains two separately determined orbits w.r.t. ICC7 and

ICC9, respectively. This is due to the mentioned bias effects (e.g., 3 frames are only suitable to determine a support plane). Further, wrongly united inf-files that do not describe the same meteor appearance, can be sorted out that way.

All files are stored in an SQLite database, a script cleans all wrongly determined orbits and provides a database with a set of meteor dynamics. For each meteor one of the two solutions is preferred that has a smaller determined error (velocity determination). Also, as a pre-filter, no solutions are considered of meteors with 3 frames. A full description of the content of the database can be found in the appendix B in table B.2.

4.4.2. Database & Content

This section describes the database and its content. To give a detailed and sophisticated overview, first, a few technical aspects are described. The velocity determination performance is shown, the filtering effect on the photometric centres with the RANSAC algorithm and further measures. Additionally, an overview of the scientific content is given.

A complete database that is described and used in this work can be obtained directly from the author. A copy can be found on the *idust* server (filename: *cilbo_science_cleaned.db*). The database contains in total 12,045 meteor orbit data. Each orbit is computed using a Monte-Carlo sampling of 128. Previous feasibility tests show that such a small sampling number is sufficient and a good compromise between statistical quality and computation performance on a personal computer.

Video Frame Un-Biasing

As already mentioned, a RANSAC algorithm is applied on the photometric centres. This is done to remove highly erratic measurements or false data as e.g., photometric centres of nearby stars. To determine how the RANSAC performes, the biased and unbiased (RANSAC applied) data are used. Figure 4.30 and 4.31 show the effects of the used method on the ICC7 and ICC9 data, respectively. Both figures show a 2 dimensional histogram. The X-axis indicates the number of video frames stored in an inf-file. The Y-Axis indicates the number of removed video frames from this data set. A colour code indicates the percentage distribution along a column. E.g., for ICC7: For around 10 % of 6 video frame long meteors 3 frames are removed.

The ICC9 camera data are slightly more affected by the RANSAC than ICC7. Relatively, more ICC9 data are filtered and more short frame appearances are affected. This can be explained by the dwell time, described in section 4.3. The apparent velocities of meteors recorded with ICC9 are smaller than for ICC7, due to the pointing and smaller angular distances to the sporadic sources. The higher pixel dwell time causes a higher chance of overexposed pixels and consequently a worse accuracy, since the photometric centre cannot be determined properly. A larger deviation of the centre from the "ideal" line is then sorted out by RANSAC.

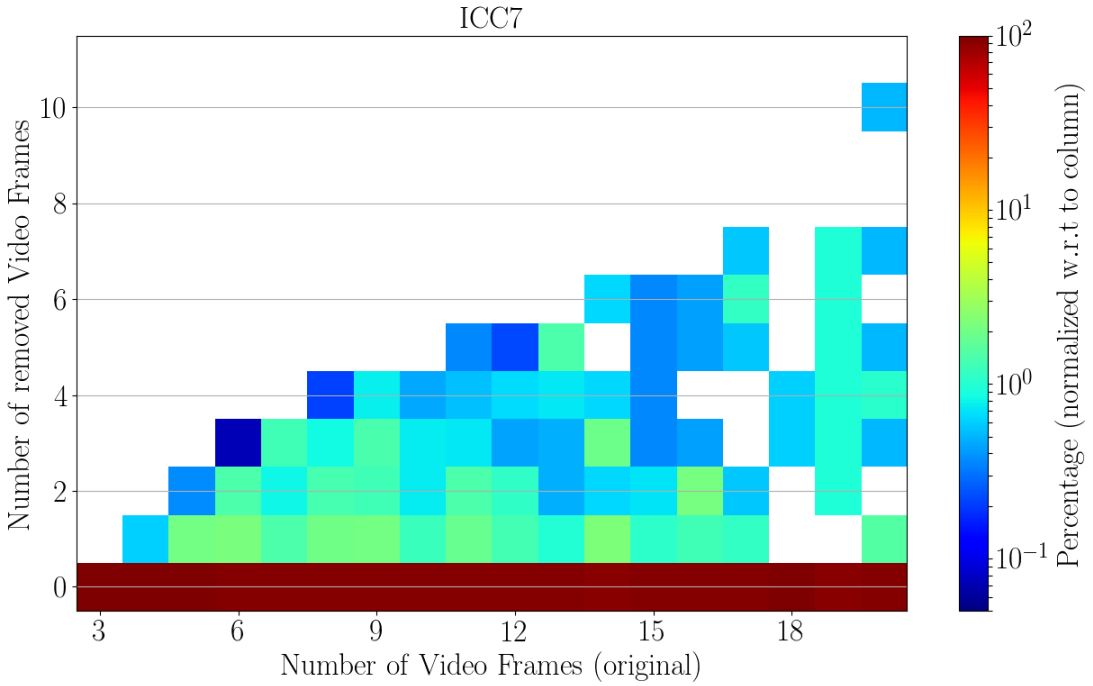


Figure 4.30.: Visualisation of Random Sample Consensus (RANSAC) filtered video frame data (Intensified CCD Camera (ICC)7). Each bin represents the percentage of filtered out video frames per total number of frames (indicated on the X-axis). The values are normalised along a column.

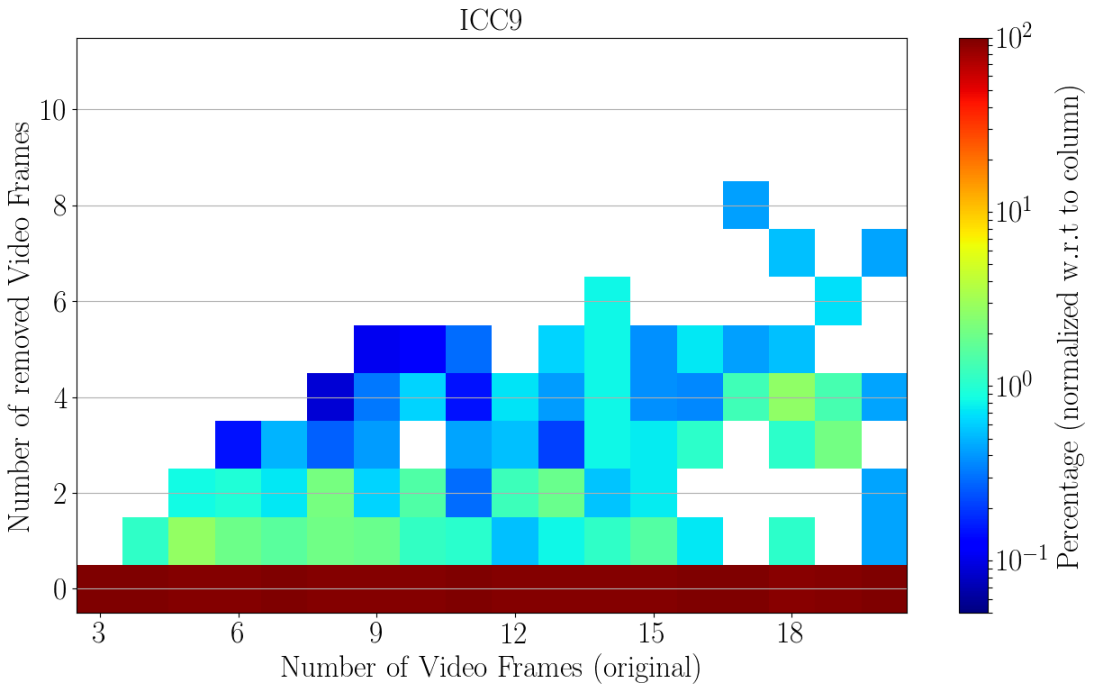


Figure 4.31.: Visualisation of Random Sample Consensus (RANSAC) filtered video frame data (Intensified CCD Camera (ICC)9). Each bin represents the percentage of filtered out video frames per total number of frames (indicated on the X-axis). The values are normalised along a column.

Fit accuracy improvement

The filtered or debiased video frame recordings lead also to an improvement of the trajectory fitting. To verify this, the database contains several parameters from the fitting procedure. An entry contains the maximum distance of a determined point to the fitted trajectory line. This maximum value is stored for the biased and unbiased data set.

These two data sets are extracted from the database and a KDE is applied, to determine a smooth distribution of the maximum deviations. Figure 4.32 shows the resulting normalised density distributions for the biased (blue) and unbiased (red) sets for ICC9. The deviation is shown on the X-axis in meters and ranges from 0 to 500 meters. For better readability, larger deviations are not shown. Compared with the biased data set, the unbiased distribution appears less wide and is further shifted towards smaller deviations. Both functions however, are skewed towards larger values and do not appear symmetrically. The median and corresponding IQR of the deviation $\tilde{d}$ for both sets are, where b and ub represent the biased and unbiased set respectively: $\tilde{d}_{\mathrm{ICC9,b}} = 194^{+104}_{-60}$ m and $\tilde{d}_{\mathrm{ICC9,ub}} = 129^{+67}_{-39}$ m with corresponding IQR widths of 164 and 104 meters, respectively.

The RANSAC algorithm improves the trajectory computation, by removing data points with large deviations. The median of the maximum deviation improves by around 60 meters and also the 50 % solution space around the median narrows by around 30 %. In total, the trajectory quality improves, leading to a more accurate orbit determination.

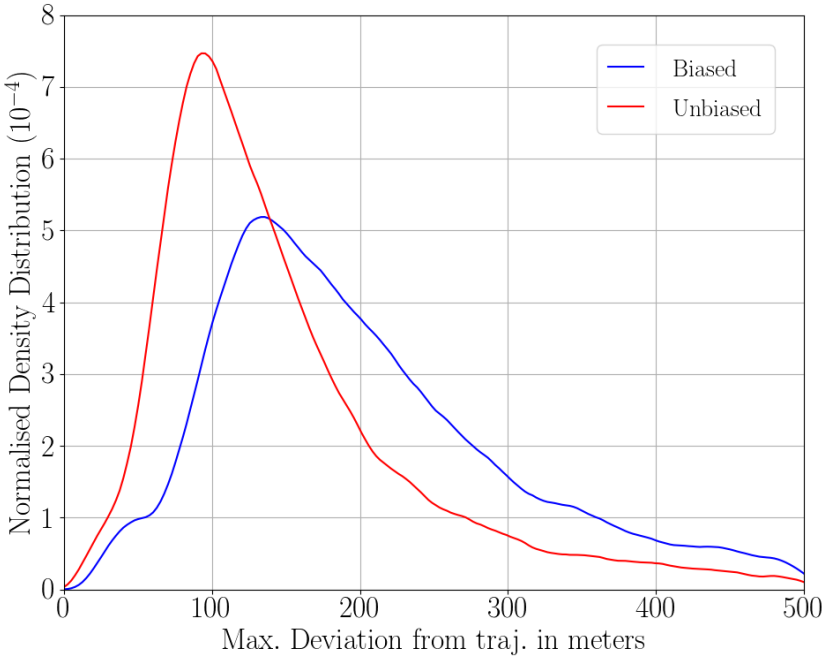


Figure 4.32.: Normalised density distribution of the biased (blue) and unbiased (red) maximum deviation data. Two Kernel Density Estimator (KDE)s are used to smooth the data. The deviations are given in meters and range from 0 to 500 meters. Larger deviations are not displayed.

Velocity measurement performance

As already described, several effects cause biases in the velocity determination. Several options are available to compute an entry velocity of a meteor. The database, as it is described here, fits a linear or quadratic function on the data sets. Further, the chosen method neglects the first and last values and extrapolates the meteor's velocity to its first appearance. Here, the resulting velocities and corresponding errors are shown as well as the effects on the scientific data.

To determine the velocity determination performance, the median values with their MAD value are extracted from the database. These data are shown as a white dotted scatter plot in figure 4.33, where the MAD is shown vs. the median; both given in km/s. Both statistical parameters result from the MC computation and each dot represents one meteor. The data describe two clouds for rather slow (less than 45 km/s entry velocity) and rather fast (faster than 45 km/s entry velocity) meteors. These separate "clouds" are associated with the average entry velocity of stream and sporadic meteors as mentioned in Albin et al. [2015a].

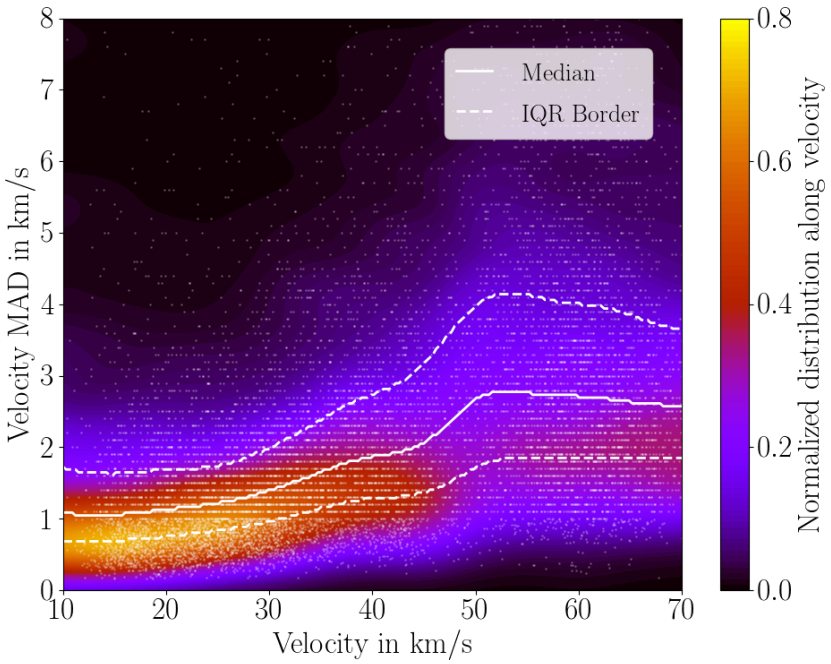


Figure 4.33.: Density plot of the velocity Median Absolute Deviation (MAD) vs. the median velocity, both given in km/s. The white scatter dots represent actual data from the orbit database. A Gaussian based Kernel Density Estimator (KDE) is applied on the data. The resulting density plot is then normalised along the entry velocity (along a column). Thus, a "data slice" along an entry velocity is a normalised density distribution with an integral of 1. A colour code indicates the density. For a proper visualisation, the median and Inter Quartile Range (IQR) boundaries for each sliced distribution is computed and drawn as a solid and two dashed lines, respectively.

A sufficient analysis of the scattered data can be performed by a column-normalised KDE. First, a KDE is applied on all data. The resulting distribution describes a normalised density plot for the entire domain. Then, the column values are normalised along the median velocity. Thus, e.g., the MAD density distribution along the median velocity 50 km/s is integrated 1.

Further, to simplify the interpretation of the resulting distributions, a solid and two dashed lines describe the median and corresponding IQR limits for each column-wise distribution. For example, the average MAD for an entry velocity of 50 km/s is around 2.8 km/s. The 50 % of the MADs lay in a range between 1.7 and 3.9 km/s.

CILBO's velocity determination performs very well up to 20 km/s. The MAD is around 1 km/s. 50 % of the data lay in a range between 0.7 and 1.7 km/s. Afterwards, the MAD increases and also the possible solution space increases, too. Especially sporadic meteors with an entry velocity larger than 50 km/s perform worse. The MAD is between 2.5 and 3.0 km/s. A larger MAD corresponds to a larger solution space for velocity depending orbital elements.

Statistical overview

Over 12,000 meteor orbits are in the database. It is expected that sporadics as well as shower meteors are in the database for further scientific analysis. Here, a quick overview of the science data is given. During the development of the Monte-Carlo extension code, it was found that

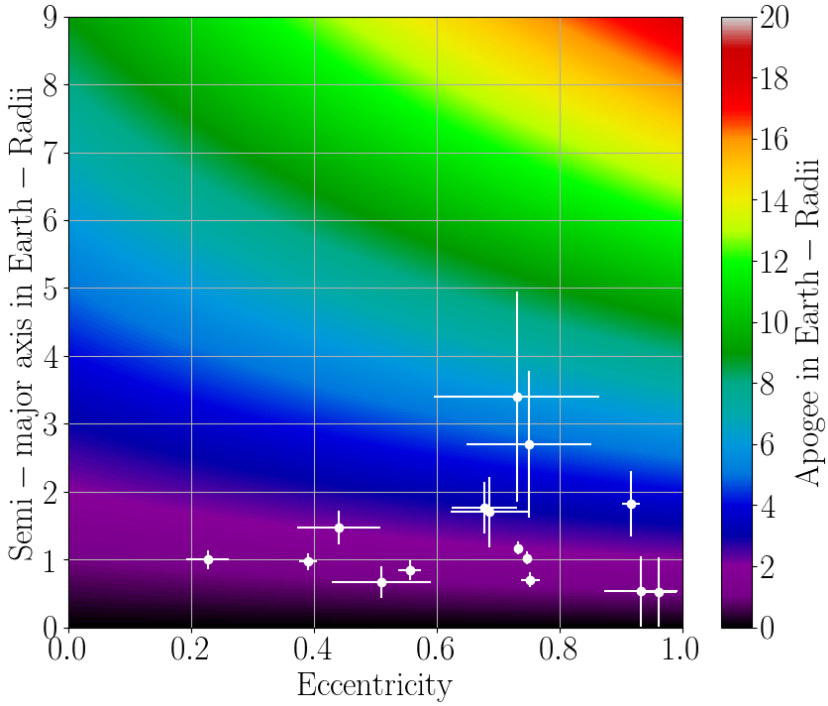


Figure 4.34.: Semi-major axis vs. eccentricity w.r.t. Earth of possible Earth-bound particles, detected by the Canary Island Long-Baseline Observatory (CILBO) system. The dot and corresponding lines indicate the median and Median Absolute Deviation (MAD) values. A colour map shows the apogee depending on the X- and Y-axis.

in rare cases, the eccentricity w.r.t. Earth is smaller than 1. This indicates an Earth-bound orbit. The current database contains 15 meteors, with a confidence larger than 95 % that are associated with this kind of orbit. Figure 4.34 shows a scatter plot of the semi-major axis vs. eccentricity with the corresponding MAD bars. A colour map indicates the Apogee in Earth-Radii. The eccentricity varies between 0.2 and almost 1.0. Whether these particles

can be associated with space debris or "mini-moons" requires further investigation. Work on the actual raw data is necessary to determine whether particle populations are measurable, or whether more research needs to be done for velocity unbiasing.

989 meteors have an eccentricity larger than 1.0 in helio-centric coordinates, corresponding to Sun-unbound or interstellar meteors. The fraction of these candidates appears rather larger, however most particles of this subset have an eccentricity between 1 and 2. This could be associated with a worse velocity determination, and partly, with actual un-bound meteors. 11 meteors have an eccentricity larger than 2. 2 meteors appear with more than $e > 10$. These extreme cases are investigated in detail. It appears that one meteor is a "mis-match" of two inf-files describing two different meteors at approximately the same time. The second meteor could not be determined properly. The velocity increases from 100 km/s to 150 km/s, indicating that no proper data are available for this case.

The remaining particles provide suitable scientific data for further analysis. Strong velocity depending parameters, like e.g., the eccentricity or argument of periapsis show, as expected, a strong correlation with the velocity uncertainties for higher entry speeds. Figure 4.35 shows a scatter plot of the eccentricity MAD vs. the median eccentricity. Most recorded meteors are rather eccentric and only a few have a $e < 0.4$. The corresponding errors appear for most entering particles suitable. Even high eccentric meteors with $e > 0.6$ have a corresponding MAD mostly smaller than 0.1. Since the velocity measurements are the most uncertain ones, these results appear feasible for further scientific studies; the average MAD value is 0.057.

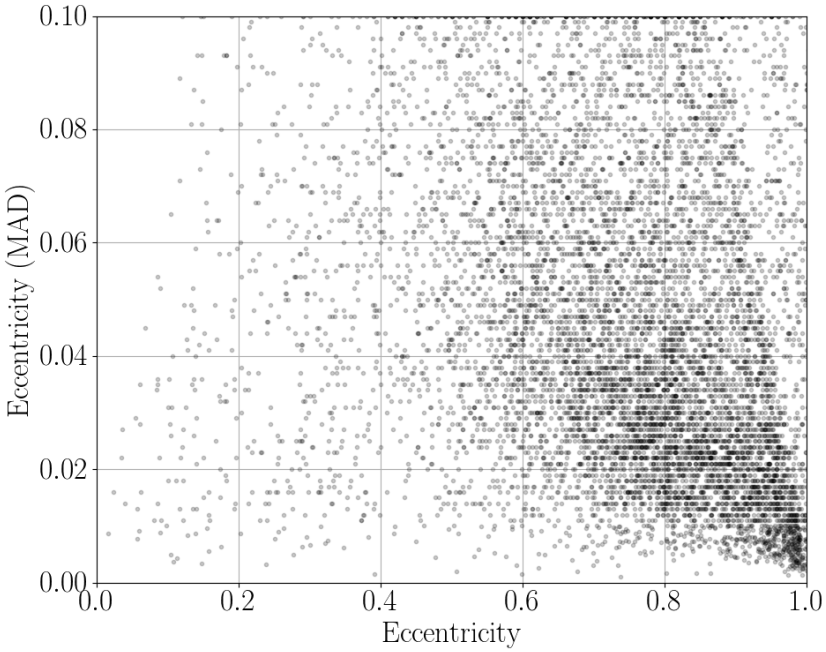


Figure 4.35.: Scatter plot of the eccentricity Median Absolute Deviation (MAD) vs. the median eccentricity of detected interplanetary dust particles.

The complete MAD statistics can be found in table 4.2. The first column indicates the lower and upper IQR boundary, indicated with the 25 % and 75 %, respectively. 50 % correspond

to the median. Most MADs have suitable values. One parameter that depends more on the astrometric accuracy is the inclination. The average MAD value is only 1.4 degrees. Only the semi-major axis has an average MAD of around 0.5 AU. This appears, compared to the feasibility of the other parameters, rather large and is probably linked to a high periapsis uncertainty.

Table 4.2.: Median Absolute Deviation (MAD) statistics of the determined interplanetary particles. The first column indicates the lower (25 %) and upper (75 %) Inter Quartile Range (IQR) boundaries as well as the median (50 %) of the Median Absolute Deviation (MAD) distributions, taken from the orbit database. The second to last column show the corresponding Median Absolute Deviation (MAD) statistics for the semi-major axis in AU, the eccentricity, the inclination in degrees, and the longitude of ascending node and argument of periapsis in degrees, too. E.g. the median Median Absolute Deviation (MAD) value of the determined eccentricities is 0.057.

Boundary	a/AU	e	i/degree	Ω/degree	ω/degree
25 %	0.22	0.028	0.69	0.00013	0.79
50 %	0.54	0.057	1.40	0.00045	2.10
75 %	1.70	0.110	2.60	0.00170	6.20

Further, to prove the suitability of the orbit computation code an analysis on the Tisserand parameter and radiants are performed, to verify the dynamic variety and correct trajectory computation by identifying well known sporadic and stream sources.

A Tisserand parameter related analysis on meteor data was performed by Subasinghe et al. [2016] with the Canadian Automated Meteor Observatory. They analysed especially fainter meteors with apparent magnitudes larger than 3. Their determined ratio of Tisserand parameters is: 49 % Damoclid-, 14 % Jupiter Family Comet- and 37 % Asteroid-related orbits. To compare the results with the CILBO dataset, the Tisserand median and corresponding error are used. Based on the error, each Tisserand value is replaced by an individually wide KDE. The resulting normalised density distribution, clipped in the range 0 to 5, can be seen in figure 4.36. Colour filled areas under the distribution function indicate the individual relations or dynamical populations, respectively. For the complete dataset, the following percentages are present: 25 % Damoclid-, 25 % Jupiter Family Comet- and 50 % Asteroid-related orbits. Filtering for fainter meteors (larger 3 mag), the percentages become: 19 % Damoclid-, 24 % Jupiter Family Comet- and 57 % Asteroid-related orbits. Subasinghe et al. [2016] detected 2 times more Damoclid related orbits, than CILBO. Only the fraction for Jupiter Family Comets related orbits appears to be similar. This large deviation could be caused by different pointing directions of the camera systems. As mentioned, CILBO points towards East and West, while the Canadian system, consisting of two cameras, points towards the North (locations: Tavistock and Elginfield, both in Ontario, Weryk et al. [2013]). As described in the previous sections regarding the technical analysis, different pointing directions have a severe effect on the detectability of meteors. Whether this effect causes a shift in the dynamical distribution needs further investigation. If so, unbiased meteor statistics can only be achieved by merging

results from different camera systems. Thus the CILBO results would only represent a part of entering dust particles.

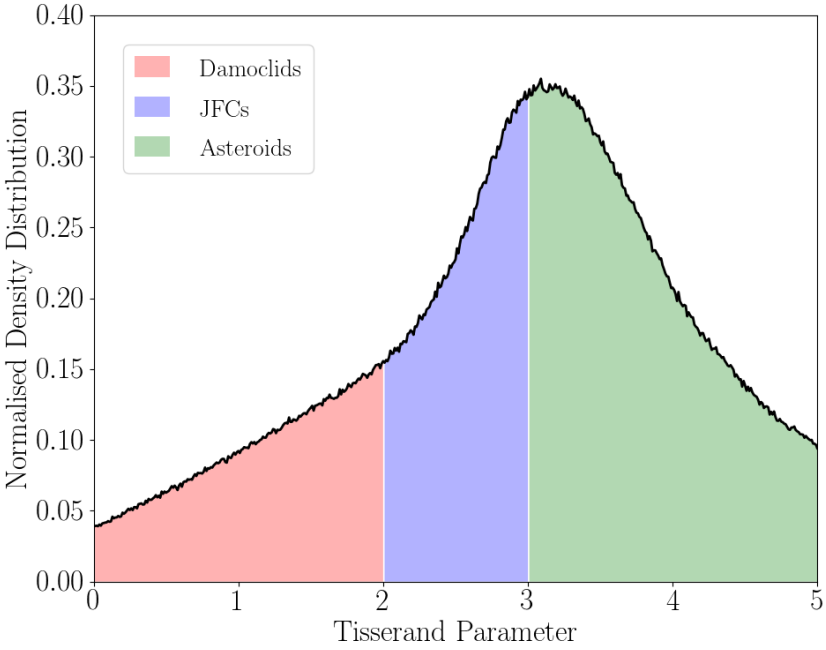


Figure 4.36.: Normalised density distribution (based on a Kernel Density Estimator (KDE)) of all Tisserand parameters that correspond to interplanetary particles. On each data point a Kernel Density Estimator (KDE) was applied, depending on the accuracy of the value. For better readability, the graph is clipped between 0 and 5. Furthermore, as a guideline, colours indicate the different Tisserand-depending populations: Damoclids, Jupiter Family Comets (JFCs) and Asteroids.

A final verification of the computed dynamics of the recorded meteors are the Earth-centred right-ascension / declination, and SOI related Earth-centric rotating coordinate system (ECRC) radiant maps as shown in figure 4.37 and 4.38, respectively.

Figure 4.37 plots the computed median radiants and colour-codes the entry depending on the entry velocity. Randomly scattered radiants are present, as well as radiant clusters, with the same velocity. These clusters correspond to meteor streams. A top and bottom aligned white triangle indicate the theoretically expected radiants of the Perseid and Geminid shower, respectively. Data points cluster in the vicinity of the drawn radiants. The expected median velocities of 58 km/s for the Perseids and around 35 km/s for the Geminids correspond perfectly with the computed data sets (Note: both meteor showers are taken as an example and do not correlate in a way).

Besides the meteor streams, sporadic meteors can be seen in an RA / DEC plot as randomly scattered radiants. However, as mentioned in the astrodynamics chapter, a proper coordinate system reveals the sporadic background sources. This ECRC radiant map is shown in figure 4.38. Here, the position of the Sun is indicated as a yellow dot at 0 degrees longitude. For a proper and less crowded visualisation, all data points have a transparency of 50 %. These computed radiants at Earth's SOI unbias the Zenith attraction effect, since Earth gravitational pull is considered. At around 270 degrees longitude is a cluster with entry speeds between 50 and

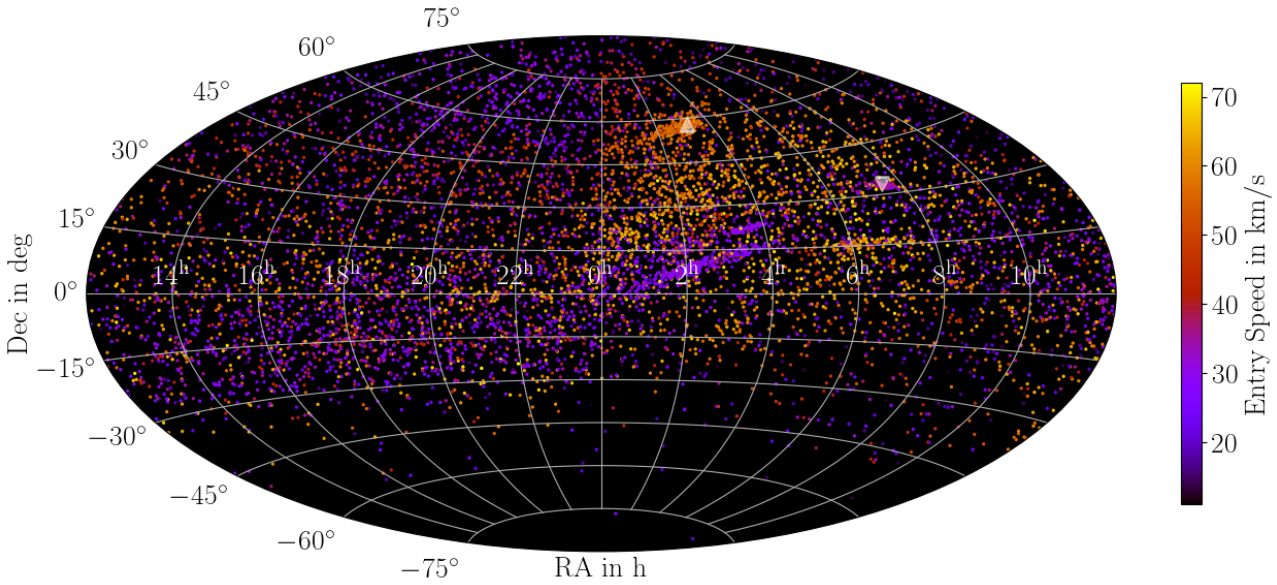


Figure 4.37.: Radiant map of the recorded meteors in equatorial J2000 coordinates, as seen from Earth. The right ascension is given in hours and the declination is given in degrees. A top and bottom aligned triangle indicate the expected average radiants for the Perseid and Geminid showers. The entry velocities are colour coded and given in km/s.

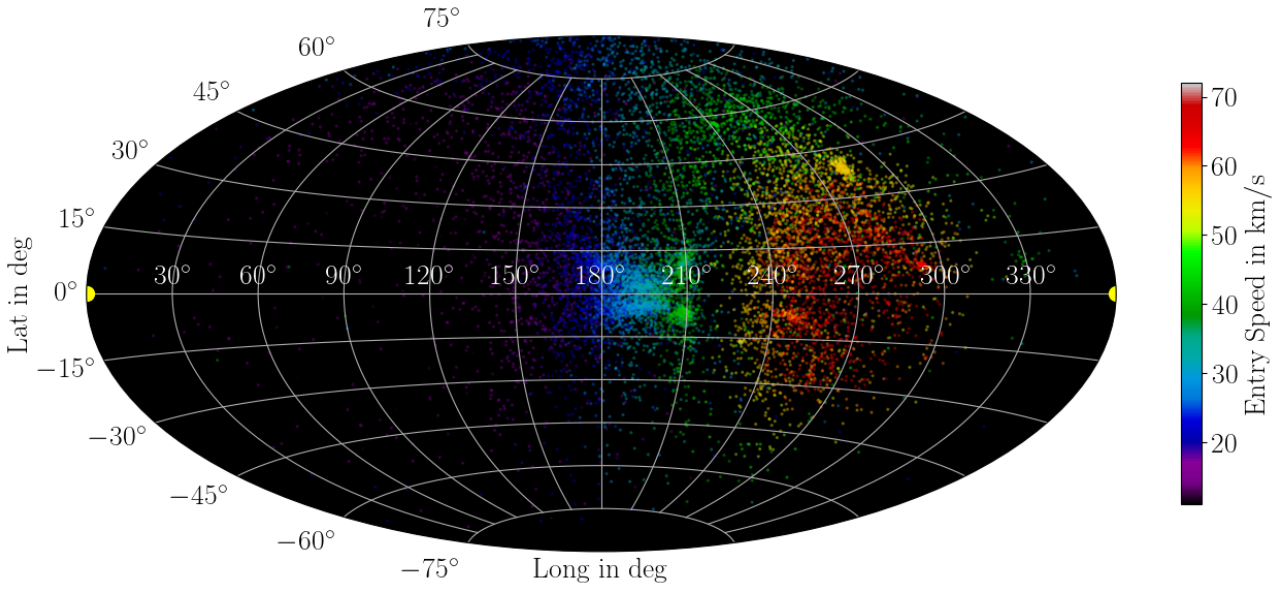


Figure 4.38.: Radiant map of the recorded meteors in Earth-centric rotating coordinate system (ECRC), as seen from Earth. Longitude and latitude are given in degrees. The entry velocities are colour coded and given in km/s. A yellow circle indicates the direction of the Sun. For better readability and identification of the northern and southern Apex source, a transparency filter of 50 % is applied on the data points.

70 km/s. The cluster is slightly divided into a northern and southern part that correspond to the northern and southern Apex source, respectively. A second source is the Antihelion source, at around 180 degrees longitude. The entry speeds are, as expected, smaller. A Toroidal source cannot be identified clearly and could be investigated in future research work. Several smaller and denser clusters indicate again miscellaneous stream meteors. Meteors from the Helion direction are not present and can only be detected by daytime operated radio observatories.

Anti-Apex meteors are also not present. Since these meteors are associated with very slow entry speeds the detectability of this population is more difficult compared with Apex meteors.

4.4.3. Conclusion / Lessons learned

In this section the Monte-Carlo based extension was shown and explained that is currently implemented in the official Meteor Research Group's MOTS routines and tools. Based on the derived biases, especially the velocity determination bias, the Monte-Carlo code computes feasible orbital elements and other dynamic parameters of stereoscopic observations. Also, a RANSAC algorithm improved the data even further by neglecting possible wrongly determined photometric centres.

All data are stored in an SQLite database. Further, for future research, this database was cleaned to provide a feasible dataset. However, miscellaneous routines for data cleaning probably did not remove all wrongly determined meteor orbits. Very few entries still have extremely high eccentricities and / or are linked to trajectories with velocities that are not scientifically correct (e.g., a meteor with an entering speed of 200 km/s that accelerate).

The section described also the resulting distributions, especially the errors (MAD) and gave an overview of the scientific data. In general the database is suitable for further scientific work. Meteor stream showers, like the Perseids or Geminids were identified, as well as sporadic sources like the Apex, or Antihelion source. This verifies the suitability of the orbit computation code. Also, the corresponding entry speeds fit perfectly.

To identify streams or to link meteors with possible parent bodies, like e.g., comets or asteroids, a second database would be recommended that contains e.g., the D_{SH} value of all meteors with all known minor bodies. Also a more detailed analysis and un-biasing of exogenic meteor candidates (eccentricity larger than 1), would be recommended, to verify observations performed by Baggaley [2000] and Baggaley et al. [2007].

4.5. Conclusion and Outlook

In this chapter the CILBO was shown. An overview was provided about the scientific tasks, its configuration and data reduction pipeline. With miscellaneous data analysis methods, like KDEs, a detailed analysis of the technical biases was provided that could also be applied on other stereoscopically observing meteor stations. Biases that affect e.g., the astrometric accuracy or a proper velocity determination were used to develop a Monte-Carlo based orbit computations extension for MOTS. The resulting database contains over 12,000 orbits and an overview of the data was provided to prove that the code works properly. Sufficiently small errors of the astrodynamical properties will lead to proper future scientific work. Furthermore, the developed software is currently (2018) still under development and will be implemented in the set of tools of the Meteor Research Group. These tools will be publicly available.

Several tools, methods and developments that were performed for the meteor research shown in this chapter are used for further technical and scientific analysis of dust measurements performed by Cassini's Cosmic-Dust-Analyzer (CDA) as will be shown in the next chapter.

5. Application: The Cassini CDA

This chapter describes the Monte-Carlo (MC) and Machine Learning (ML) based projects on Cassini's in-situ instrument Cosmic-Dust-Analyzer (CDA). CDA operated in the Saturnian system from 2004 to Cassini's final plunge in September 2017. During that time the dust instrument recorded millions of dust particles and hundred-thousands of mass spectra.

The first major data scientific approach is the determination of the astro-dynamical properties of the impacting dust particles. A MC based orbit computation procedure is shown that uses miscellaneous calibration functions and intrinsic geometric properties of the instrument. Based on this, a recent update is shown on the interstellar direction based on Interstellar Direction (ISD) particle candidates by Altobelli et al. [2016]. Then a section shows a new Neural Network (NN)-based approach to re-calibrate the velocity functions determined by Srama et al. [2004]. Additionally, a first function is provided for Wall-impacts. The last part uses differently developed ML algorithms [Burkhardt, 2017, Knieling, 2017] to find unknown type 4 spectra in the science database[1].

The projects are done mostly in parallel. Thus, results from one section influence the work and projects of other ones. However each section can be read and understand individually. Example: During the ML based project, the already labelled data have been astro-dynamically analysed. The analysis has revealed that most particles have been detected in the vicinity of its periapsis. The corresponding dwell time is rather short (e.g., compared with the apoapsis phase) and thus, the probability of detecting particles in their periapsis is small. However, the "periapsis-near" detection are over-represented in the data set, indicating that the detection or velocity determination is biased. This leads in parallel to the NN-based velocity calibration project.

5.1. Introduction

The CDA is a dust measuring device on the Cassini Orbiter and has mainly been developed in the Max-Planck Institut für Kernphysik in Heidelberg, Germany. Its mounting position on the Cassini orbiter and an actual image of the dust detector can be seen in figure 5.1 a) and b), respectively. Since the device is mounted on a revolvable table CDA's pointing direction,

[1]Type 4 dust particles are rarely detected dust particles with no water-ice. These mineral dust particles are associated with retrograde dynamics from the moon Phoebe or originate from interplanetary or interstellar space.

hereafter called boresight, is not fixed as shown in figure 5.1 a). A rotation of 270° allows a boresight optimization depending on the scientific objective like e.g., close pointing to the Dust- or Plasma-RAM, ISD or Interplanetary Dust Population (IDP) direction or during plume fly-throughs. A list of all main scientific objectives is shown in table 5.1.

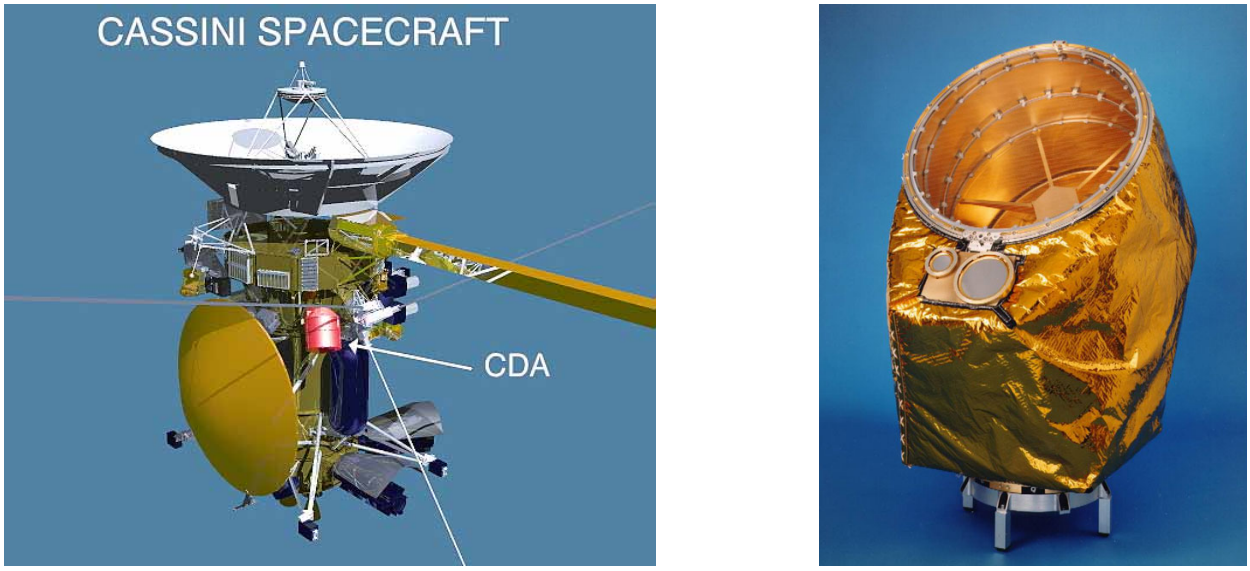


Figure 5.1.: (a) Cosmic-Dust-Analyzer's mounting position on the Cassini orbiter. Due to the rotation table, the boresight of the Cosmic-Dust-Analyzer is not fixed. Here, Cassini's configuration is shown before 2005, where the Huygens probe has not been separated yet. (b) An image of the final Cosmic-Dust-Analyzer setup with Multilayer Insulation. The Cosmic-Dust-Analyzer consists of two sub-systems, the Dust-Analyzer and the High Rate Detector. The High Rate Detector consists of two small polyvinylidene fluoride sensors which can be seen beneath the opening tube of the Dust-Analyzer. The entrance of the Dust-Analyzer is 0.41 m in diameter. The inclined grids of the QP channel can be seen as well as the front view of the flight-mass-spectrometer in the center. The revolvable table can be seen on the bottom of the image. Image Credit: [NASA/JPL, a] and [Srama et al., 2004]

Table 5.1.: Cosmic-Dust-Analyzer (CDA)'s science objectives. From Srama et al. [2004].

Cruise science
– Extend studies of interplanetary dust to the orbit of Saturn.
– Sample the chemical composition of dust in interplanetary space and across the asteroid belt.
– Determine the flux of interstellar particles during solar maximum conditions.
– Search for dust streams originating from Saturn.
Jupiter flyby
– Investigate the dynamics of the Io dust streams as discovered by Ulysses and Galileo. Characterize their direction, size-mass-distribution and correlation with the jovian and interplanetary magnetic field.
– Investigate the dust stream fluxes caused by the jovian system with respect to the Jupiter distance.

– Analyze dust stream particles at a different epoch from Galileo.

– Characterize the elemental composition of dust stream particles.

Rings

– Map size distribution of ring material, search for ring particles beyond the known E-ring.

– Analyze the chemical composition of ring particles.

– Study dynamical processes (erosional and electromagnetic) responsible for the E-ring structure, study interactions between the E-ring and Saturn's magnetosphere, search for electromagnetic resonances.

– Determine dust and meteoroid distribution both in the vicinity of the rings and in interplanetary space.

Icy satellites

– Define the role of meteoroid impacts as mechanism of surface modifications.

– Obtain information on the chemical composition of satellites from the analysis of gravitationally bound ejecta particles in the vicinity of the satellites (within Hill spheres).

– Investigate interactions with the ring system and determine the importance of the various satellites as a source for ring particles.

Magnetosphere of Saturn

– Determine the role that dust plays as source and sink of charged particles in the magnetosphere.

– Search for electromagnetically dominated dust (small particles) and for dust streams.

CDA consists of two sub-systems, the High Rate Detector and the Dust-Analyzer (DA). Two polyvinylidene fluoride foils are the main sensor of the High Rate Detector device and can be seen in figure 5.1 b). Two silverly round plates beneath the tube opening are the foils. Due to the one second dead time of the DA the High Rate Detector is used to determine the flux especially in high dense regions, like during an E-Ring crossing. The second device, the DA, is the tube-like device shown in figure 5.1 a). This work focuses on this sub-system. A schematic cross section of the DA is shown in figure 5.2.

CDA is has two targets with a total number of six measurement channels[2] and has an

[2]5 channels are "Q" channels; one is a multiplier. The "Q" in the name abbreviations stands for charge

opening diameter of 0.41 m. Two impacting dust particles are shown in the technical drawing to illustrate possible channel answers of the DA. First, a dust particle passes four grids, each with a transmission of 95 % where the two inner grids are inclined by 9° with respect to the outer grids. These two grids are the QP (P: Primary) channel, which allow the determination of the particle's charge down to 1 fC. During the fly-through of the first grid, the induced voltage signal increases to a certain level, as can be seen in the schematic QP curves in figure 5.2. While the particle moves within the transmission grids, the signal remains approximately constant. After passing the second grid the signal decreases to its original level. An incoming dust particle has 3 impact possibilities: wall impact, Impact Ionisation Detector (IID) impact or Chemical Analyzer Target (CAT) impact. The IID is a gold coated convexly shaped target with a diameter of 0.41 m. Impacting particles cause a signal in the IID connected QT (T: Target) channel. With a diameter of 0.16 m, the rhodium coated CAT contributes a smaller sensitive area to the DA setup. In front of the CAT is a grid with a transmission of 68 % connected to the QA (A: Acceleration) channel while the target itself is the QC (C: Chemical) channel. Additionally, the CAT is supplied with a voltage of +1 kV, while the IID is grounded.

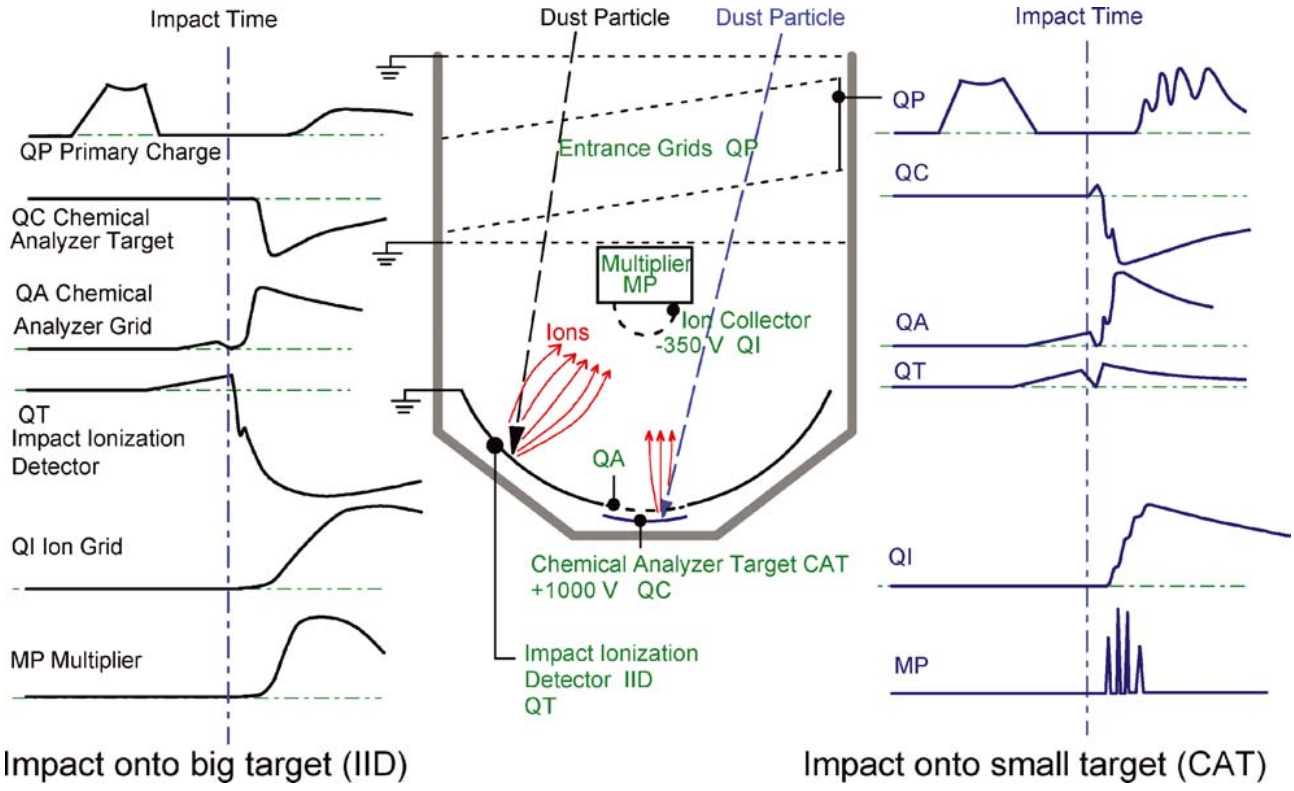


Figure 5.2.: Schematic set-up of Cosmic-Dust-Analyzer. Two exemplary impacting dust particles show the channel signals of the Dust-Analyzer for an Impact Ionisation Detector impact and Chemical Analyzer Target impact, respectively. The channel answers are shown on the left side and right side of the cross section and correspond to Impact Ionisation Detector and Chemical Analyzer Target impact. Five channels from the channels plot representatively the charge vs. the time and the multiplier channel plots the voltage vs. the time. A dashed vertical line aligns all curves w.r.t. the impact time on the corresponding target. Due to its potential difference of 1.35 kV between QC and QI, only Chemical Analyzer Target-impacting dust particles can be analyzed spectroscopically. From Srama et al. [2004].

Dust particles fragment after an impact on the IID or CAT due to their high velocity of a few km/s. The residues of the original dust particles are negatively and positively charged ions. Due to a supplied voltage of −0.35 kV on the QI (I: Ion) grid anions are accelerated to the multiplier MP which is approximately 0.23 m away from the curved target surface. Since only the QC channel is supplied with an additional voltage of +1 kV, a high resolution flight mass spectrum can only be obtained from particles impacting the CAT. This is schematically drawn in the MP signals for CAT impacts (MP signal–"comb" reveals certain mass lines), while IID impacts appear rather smoothed. Due to its large mass of 103 u the CAT rhodium target material is a suitable reference peak in the spectra since most dust particle spectra contain hydronium, sodium and other light materials. An exemplary spectrum can be seen in figure 5.3 from Day-of-year 195 in the year 2005. A detailed description of CDA and its setup, tests and its electronics can be found in Srama [2000], Srama [2009] and Srama et al. [2006].

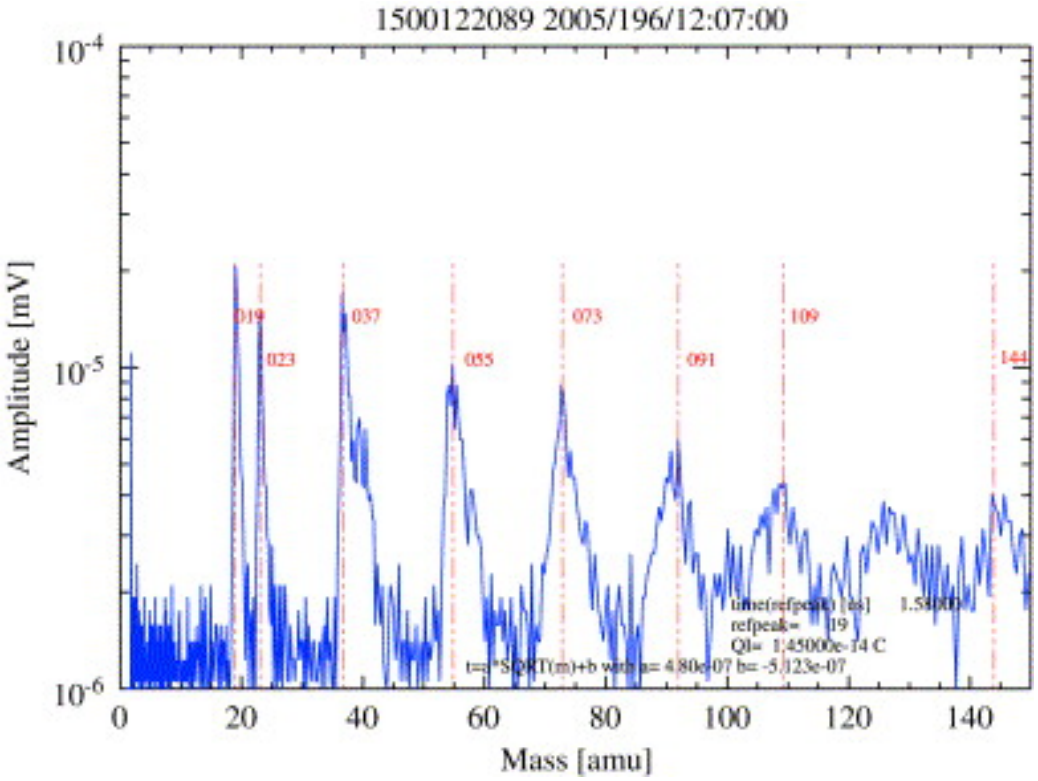


Figure 5.3.: Spectrum of an ice particle in 2005 Day-of-year 196. The amplitude is plotted vs. the atomic mass. Major peaks are indicated with red dashed lines. The peak at 19 u is hydronium H_3O^+. Since water has an atomic mass of 18 u, the peaks at 37, 55, 73, 91 and 109 show a hydronium molecule with difference water cluster sizes. From Srama et al. [2006].

CDA accomplished all of its scientific objectives. During the cruise phase and fly-by at Jupiter, a few measurement campaigns were already started. Interplanetary dust [Hillier et al., 2007, Kempf et al., 2004] as well as interstellar dust [Altobelli et al., 2003] have been detected. Also Jovian [Postberg et al., 2006] and Saturnian stream particles have been identified [Hsu et al., 2010]. Within the Saturnian system several measurement campaigns revealed Saturn's dust environment [Srama et al., 2006]. E-Ring crossings helped to determine the properties of

this extended ring [Kempf et al., 2008, Postberg et al., 2008] and close fly-bys on Enceladus have shown that the moon has a reservoir or ocean of liquid water under its icy surface [Postberg et al., 2009, Hsu et al., 2015].

5.2. Dust Orbit Computation Code (DOCC)

This section describes a new MC based approach to determine the orbital elements of CDA dust measurements in space. The shown method however can be easily applied on any in-situ dust instrument with a known angular Field-of-view (FOV), velocity measurement and pointing determination.

First, the motivation, previous works on this concept and the requirements are shown that are needed to develop a sufficient MC based orbit computation tool. Afterwards, the MC input Probability Density Function (PDF)s are explained, namely the impact directionality and velocity determination error. Following, a feasibility study determines whether, based on some artificial examples, the a-priori given PDFs deliver sufficient results. Finally, the algorithm is explained and a few test cases are shown.

5.2.1. Motivation

Chapter 3 described orbit dynamic concepts of Keplerian motion. Orbital elements, the Tisserand parameter, the Sphere Of Influence (SOI), the Hill-Sphere and many more basics were introduced that represent only a fraction of the complete topic. To develop a sophisticated orbit computation solution for CDA in-situ measurements, the scientific background of the Saturnian system needs to be considered. Based on the scientific aspects, later on, a reasoned and selected number of parameters are chosen that are computed with the DOCC. With a reasoned list of to-be-determined properties, a database can be set up that fulfills several scientific requirements for further studies.

In general, 2 different classification cases are possible that are shown separately in each section: dust grain detections within the Hill Sphere of a Saturnian moon and measurements within the SOI of Saturn but not in the vicinity of a satellite. Each case is divided in possible sub-classes, like e.g., dust orbiting a moon, or dust particles on an escape orbit leaving its parent body. Possible sub-cases are then described as sub-sections. All shown cases are implemented in the DOCC software that is described later. Scientific publications are used to determine the parameters that shall be part of the orbit computation procedure and database.

Moon solutions

Since 2004 Cassini had several moon flybys. Especially Titan that were used to change and alter the trajectory of Cassini and Enceladus were frequent targets of the spacecraft. Close flybys up to several tens of kilometers above the surface of Enceladus were done for plume

measurements in the southern hemisphere of the moon. One of Cassini's closest approaches was on 2015 Day-of-year 301, respectively October 10th. At its closest approach the spacecraft was around 45 km away from the surface as shown in figure 5.4. The solid line shows the distance and the dashed line indicates the border of the Hill-Sphere; and the X-axis shows the time with 0 seconds at closest approach (UTC 15:23).

Thus, dust that has been detected in the vicinity of Saturn's satellites can be dynamically linked with these bodies. Three possible cases are explained in the following sections.

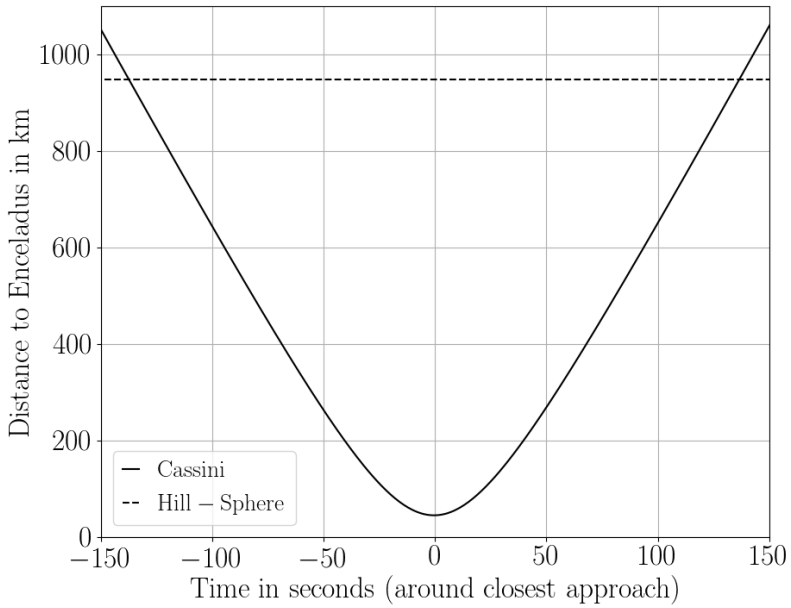


Figure 5.4.: Distance of the Cassini spacecraft to the surface of Enceladus for flyby E21 (2015 Day-od-year 301, Oct. 10th). The solid line indicates Cassini and the dashed lines shows the border of Enceladus Hill Sphere. The X-axis shows the time, normalized at the closest approach at UTC 15:23 at a distance of approximately 45 km.

Moon-bound orbit One of Cassini's major scientific objectives was the detailed analysis of the Saturnian moon-system. On November 2005, Cassini had a close Rhea flyby of approximately 500 km altitude. During this close encounter the Magnetosphere Imaging Instrument instrument measured the electron and ion flux of Saturn's magnetosphere in a range of a few tens and hundreds of keV [Jones et al., 2008]. These ions, respectively plasma co-rotates with Saturn's magnetosphere as can be seen in figure 5.5, where the directionality of the B and E field is shown in the lower right and upper left corner of the figure, respectively. The movement of an ion is indicated with an i^+ and the green "loop-shaped" trajectory indicates its path. Here, the Cassini trajectory of the flyby is shown, too. During its close encounter with the moon, the spaceraft was behind the "plasma-shadow" or wake of the moon that is shown as a dark-green area below the moon. Here, the ions flux measurement is expected to drop significantly. The yellow circularly shaped area represents Rhea's Hill Sphere with a radius of approximately 5,800 km, respectively 7.5 times the radius of Rhea (around 760 km).

The differential intensity began to drop slightly with entering the Hill Sphere of Rhea, dropped significantly in the wake and began to rise again until the spacecraft left the gravitational vicinity of the moon. Jones et al. [2008] determined that the drop cannot be explained by a thin gas exosphere of the moon but by solid particles or fragments that are captured in Rhea's gravitational field. This was interpreted as a ring or more likely a halo of dust around the moon.

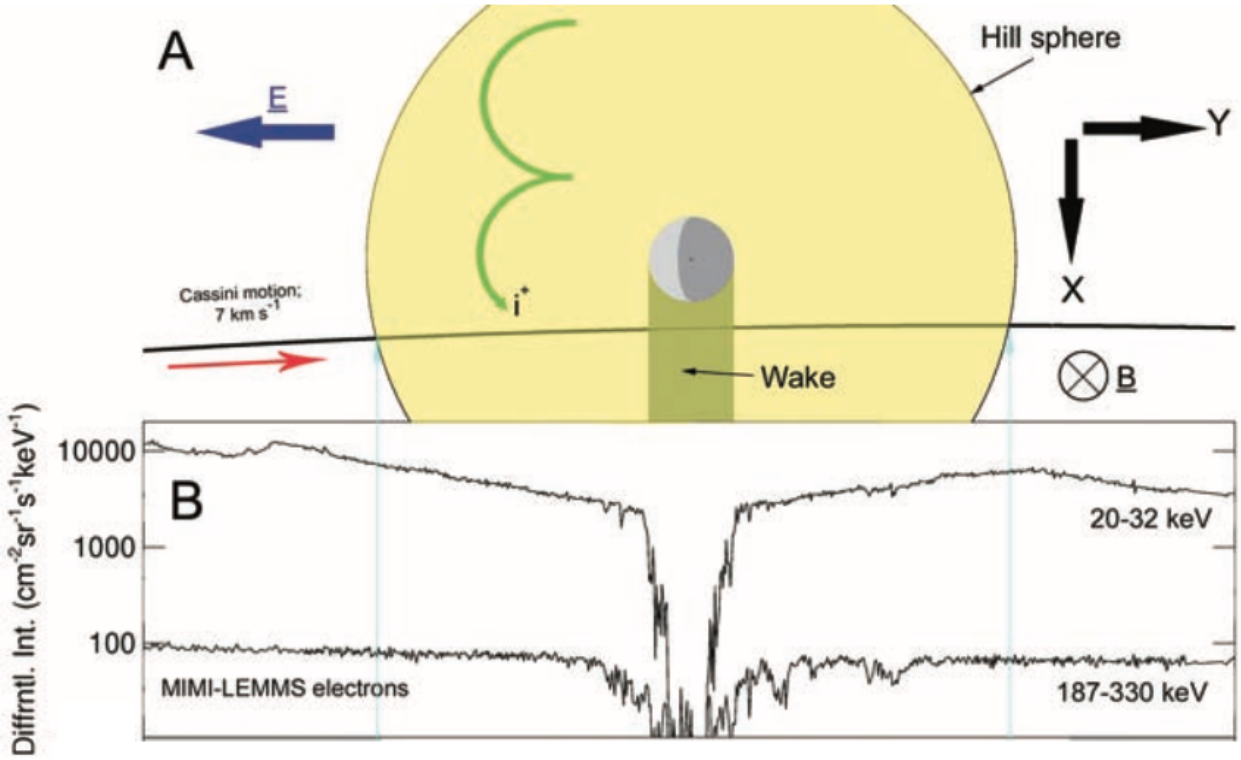


Figure 5.5.: **A** Magnetosphere Imaging Instrument measurement campaign during the close encounter with Rhea on November 2005. Rhea is shown in the center as a partly illuminated circle. The yellow area around the moon is its Hill Sphere (both, moon and Hill Sphere are in scale). Cassini's trajectory is shown as a black curve, entering the Hill Sphere and passing through the wake of the moon where no plasma or ion measurement are expected due to the coverage of the moon. Saturn's magnetic B and electric E field directions are shown in the bottom right and upper left corners of the figure, respectively. Cassini's flight direction is indicated with a red arrow. Flight directions of ions i^+ are indicated as a green loop-shaped line. **B** Corresponding differential intensity measurement of the ions in a range of $20 - 32$ keV and $187 - 330$ keV, respectively. From Jones et al. [2008].

Later, Tiscareno et al. [2010] used Cassini's Imaging Science Subsystem to observe possible rings or a dust halo with optical instruments. However, they were not successful in identifying any dust structures in the vicinity of the moon. Possibly, these structures are too faint to be observed with the Imaging Science Subsystem.

Thus, the DOCC programm shall implement possible moon-bound possibilities, to find moon-bound particles at Rhea, Dione or other moons. For example, Titan's gravitational pull and dense atmosphere could drag particles on bound orbits after close flybys. High eccentric, Titan-bound particles could collide with the moon after several close periapses and consequently high decelerations due to the atmosphere. Furthermore, Titan's Hill Sphere has a diameter of

around 70,000 km. Over 120 flybys altered the spacecraft's trajectory and several measurements campaign during these encounters could reveal such particles.

Moon-outbound orbit Saturn's ring system consists of miscellaneous rings, first discovered by G. Galilei and further investigated in Detail by the astronomers C. Huygens and G.D. Cassini. Deeper observations revealed the A, B, C and D ring in the system. With modern telescopes, in-situ measurements and observations the fainter E-Ring, Janus/Ephimetheus-Ring, G-Ring, F-Ring and Phoebe-Ring have been discovered. Especially the E-Ring is the largest Ring in the Saturnian system ranging from approximately 180 to 500 thousand kilometers. The E-Ring consists of ice particles that originate from Enceladus. An E-ring dust source was found by the Cassini spacecraft during its first months in the Saturn system. Images from the Imaging Science Subsystem instrument showed active regions in Enceladus' southern hemisphere that emit ice particles from cryo-geysers [Hansen et al., 2006, Porco et al., 2006, Spencer et al., 2006, Spahn et al., 2006]. These regions can be seen in figure 5.6 that shows the southern region of the moon with its geysers / plumes.

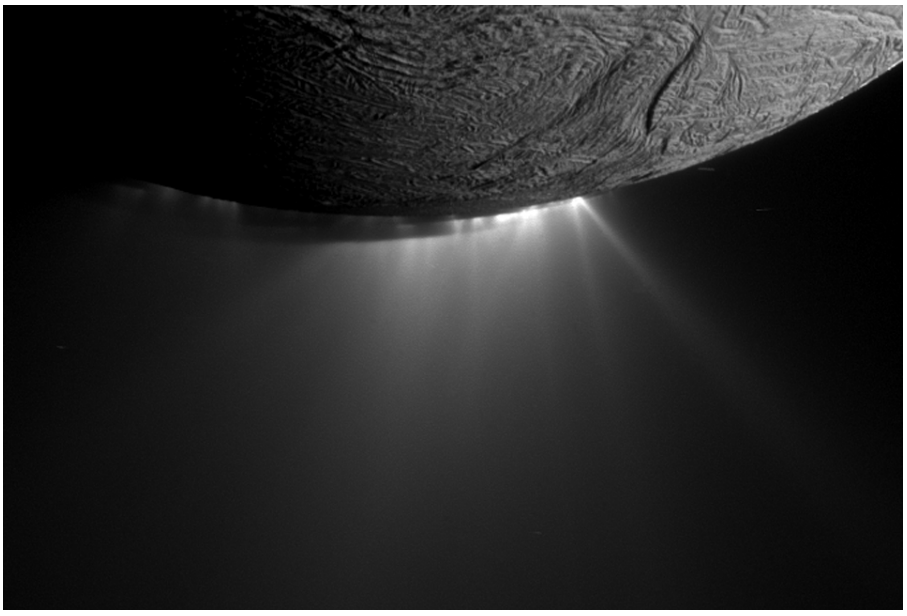


Figure 5.6.: Image of Enceladus' southern regions. This image was taken with the Imaging Science Subsystem and shows the active regions emitting dust and ice particles from the moon. Image Credit: NASA/JPLS/CASSINI/ISS

Detailed analysis of the southern regions was done by Nimmo et al. [2014] and Porco et al. [2014] who found over 100 active regions on the surface and determined the coordinates on Enceladus' surface. However, recent studies by Spitale et al. [2015] reveal that the active regions, so called "tiger-stripes" do not consist of individual geyser spots but emitting "curtains". Figure 5.7 shows a map of the active regions *Damascus*, *Baghdad*, *Cairo* and *Alexandria*. The circles indicate the individual spots found by Porco et al. [2014].

Since its arrival in the Saturnian system Cassini had several Enceladus flybys with partly close proximity to its surface. As already shown in the beginning of this section, in 2015 Day-of-year

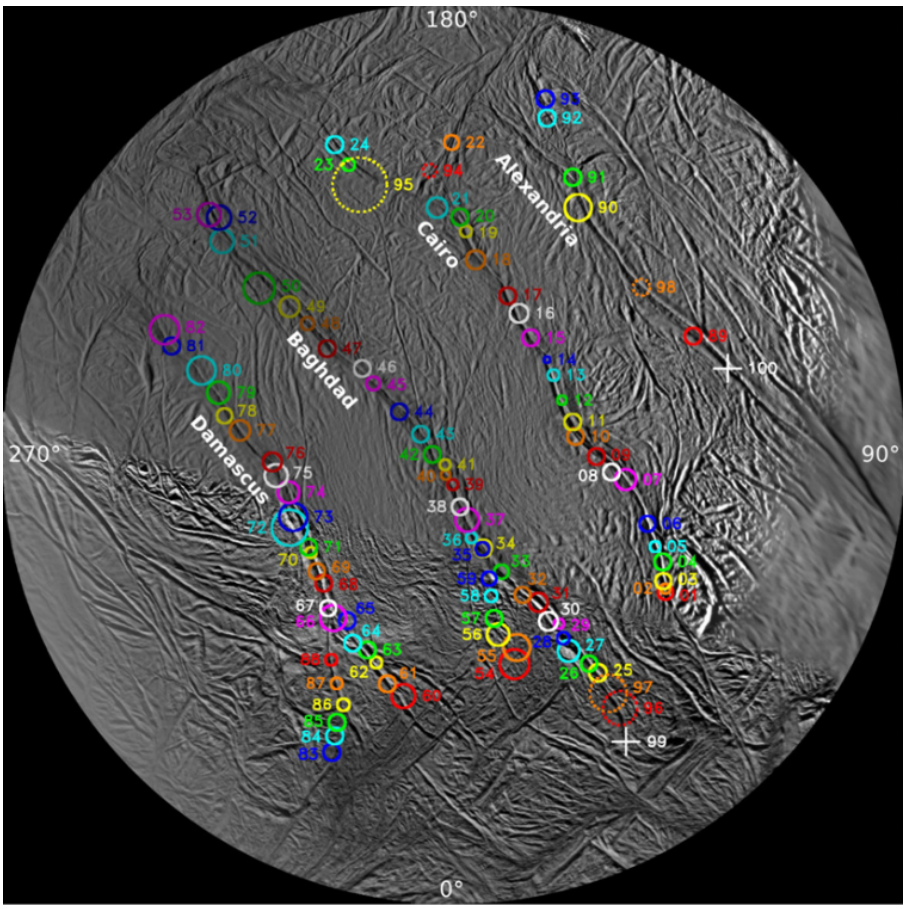


Figure 5.7.: Map of Enceladus' active regions around its south pole. The four tiger-stripes *Damascus*, *Baghdad*, *Cairo* and *Alexandria* can be seen with identified individual geyser spots shown as circles. Each spot is numbered and the corresponding coordinates and properties can be found in Porco et al. [2014]. From Porco et al. [2014].

301 Cassini flew through the water plumes in an close encounter of 45 km w.r.t. Enceladus' surface. During the several flybys and during this close encounter the CDA was online and detected thousands of dust grains. Particles escaping the moon on orbits with a moon-bound eccentricity larger than 1 could have been detected by the dust instrument. Thus, DOCC needs to identify these possible solutions for each flyby and fly-through within the Hill Sphere of all moons to find new sources or to link spectra measurements with an origin or source. Other, atmosphere-less moons like Tethys, Dione or Rhea could be hit by larger particles or meteoroids emitting ejecta that could have been detected by CDA, too. Linking spectra data with possible source coordinates would allow a compositional surface mapping of the moons.

Further moon outbound cases are flybys of dust particles. For example Dione, Tethys, Rhea or Titan are embedded in the E-Ring. Consequently, ice particles from the E-Ring can cross a moon's Hill Sphere. Depending on the proximity of the particle with the moon, its trajectory can be significantly altered. After the flyby these particles leave the Hill Sphere with different orbital parameters. These particles also appear as outbound particles. To distinguish between flyby particles and ejecta, the periapsis in a moon-centric coordinate system needs be considered. Periapsis values larger than the radius of the moon are most likely these particle flybys.

Moon-inbound orbit As mentioned, particles from the Saturnian system (e.g., E-Ring particles) can enter a moon's Hill Sphere, flyby at the moon, and leave its gravitational vicinity with altered orbital elements. Within the Hill Sphere, CDA is also able to measure these entering particles that either have a flyby or collide with the moon. A prominent example of dust deposits is the moon Iapetus [Denk et al., 2010, Spencer and Denk, 2010] that has probably a few centimeters of dust on its heading side (figure 5.8). The dark color is probably due to a organic composition of the depleted dust. Verbiscer et al. [2009] discovered the retrograde orbiting Phoebe Ring that might be a continuous source for further dust deposits.

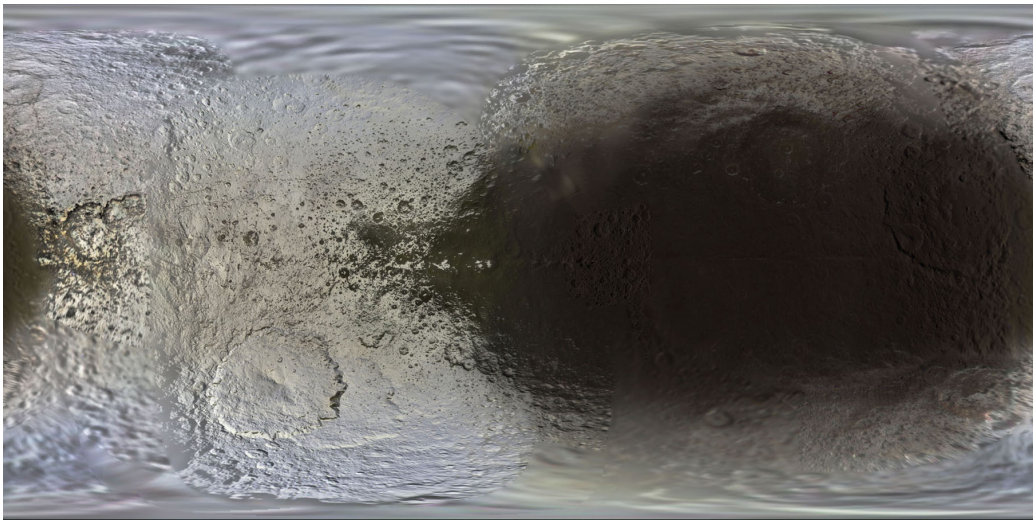


Figure 5.8.: Map of Iapetus. The trailing and heading hemisphere is on the left and right of the image, respectively. It can be seen that that Iapetus has an albedo dichotomy with a darker region in the heading side of the moon. It is assumed that exogenic and organic rich dust is mainly responsible for the darkening. Image Credit: NASA/JPL/Cassini/ISS.

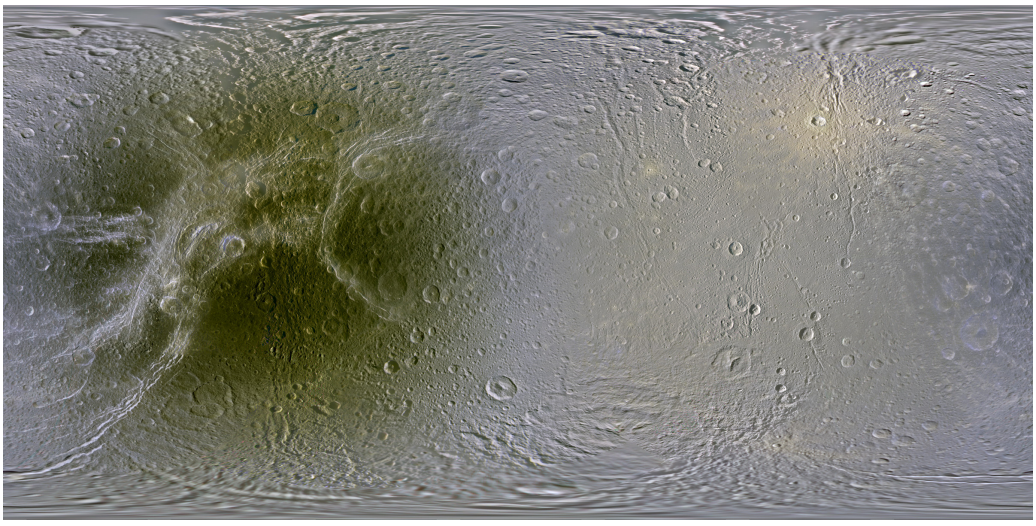


Figure 5.9.: Map of Dione. The trailing and heading hemisphere is on the left and right of the image, respectively. Due to radiation, the surface of the trailing side became darker or redder. E-Ring particles deposit on the heading side of the moon. Since these particles are mainly water ice and geologically younger than the surface of Dione, the heading hemisphere appears slightly brighter than the average surface albedo. Image Credit: NASA/JPL/Cassini/ISS.

Moons that orbit Saturn within the E-Ring show also dust deposits on their heading side, like e.g., Dione [Scipioni et al., 2013] or Rhea [Scipioni et al., 2014]. Figure 5.9 shows the surface map of Dione with the heading side on the right of the image. The heading area is slightly brighter than the average surface albedo, since E-Ring particles consist mostly of water ice and are "fresh" material compared to the surface. The trailing side (left side of the pictures) shows a darkening due to radiation effects.

Saturn solutions

Besides miscellaneous moon flybys the Cassini mission revolved Saturn mainly outside the Hill spheres fo Saturn's moons. Similarly as shown before, three dynamical cases are possible: Saturn-bound dust, and dust escaping or entering the Saturnian system.

Saturn-bound orbit During its mission the CDA determined density profiles and spectra classes in the Saturnian system, especially within the E-Ring, see for example Srama et al. [2006], Kempf et al. [2008], Postberg et al. [2008]. Beckmann [2008] analysed the dynamical properties of the E-Ring numerically. With the orbit database based on DOCC this model can be proved or denied. Thus, all orbital elements need to be computed. Furthermore, dust might be dynamically linked with the moons. Similarly as Jupiter Family Comets that are linked with Jupiter, dust might be coupled with the larger Saturnian moons. Computing the Tisserand parameter shall be thus considered.

Saturn in- and outbound orbit Particles with an eccentricity larger than 1 with respect to Saturn have an escaping or entering trajectory within Saturn's SOI. Particles entering the Saturnian system are of interplanetary or interstellar origin. A mixed case are particles from the Oort Cloud that might enter the Solar System on hyperbolic orbits but do not originate from interstellar space.

Nano streams Particles escaping the Saturnian system could result from two sources: ejecta from moons, due to a high speed impact of a meteoroid or nano dust streams. Grün et al. [1993] already discovered dust streams with the Ulysses dust detector. These particles were emitted by the moon Io in the Jovian system and were accelerated into the interplanetary space. Similar Saturnian dust streams were detected by CDA [Kempf et al., 2005, Hsu et al., 2011].

These grains are nano sized, electrostatically charged and coupled with the magnetosphere of Saturn and accelerated on an outbound orbit. These particles collide also with the moons and could be detected as moon-inbound or moon-outbound (flyby) particles. Srama [2015] measured a nano dust stream "occultation" by Titan. These occultations can be used to determine atmospheric properties of the moon, since the density and consequently drag decreases for higher altitudes.

However, due to their size and impact velocity nano grains are classified as noise by the CDA on-board electronics. Their actual mass spectra can only be determined by cross-correlating or averaging hundreds of signals, assuming a similar mass composition.

Interplanetary and Interstellar Dust During its cruise phase and in the Saturnian system, CDA detected interplanetary as well as interstellar dust [Altobelli et al., 2003, Kempf et al., 2004, Hillier et al., 2007]. Recent ISD analysis revealed 36 particles from the Local Interstellar Cloud [Altobelli et al., 2016]. These ISD candidates were detected with the CAT and showed similar mineral compositions.

IDPs and ISD particles appear as inbound as well as outbound particles that had a flyby at Saturn. Both cases need to be separated into two tables containing heliocentric bound orbits, and heliocentric un-bound orbits. Furthermore, to analyse the directionality and flux into the Saturnian system, radiant information need to be determined too, similarly as described in the Canary Island Long-Baseline Observatory (CILBO) chapter 4. Using similar coordinate systems for Saturn could reveal stream or sporadic sources, too. Furthermore, in meteor science, the entering velocity is used too for shower classification. This parameter, as well as the directions are computed at the border of Saturn's SOI. Using clustering wavelet classification methods as described in Brown et al. [2010] could link Earth's based meteor science with in-situ dust measurements in the Saturnian system. At the end of the DOCC section, an overview of a scientific outlook is given. Bötsch [2018] and Faber [2018] use new graph-based methods and dynamical properties from meteor science, to identify possible particles from interplanetary and interstellar space, respectively.

5.2.2. Previous in-situ orbit computations

In the past, two similarly built dust detectors have been launched into space, the Ulysses dust detector and the Dust Detection System [Goeller and Grün, 1989, Grün et al., 1992, 1995]. A detailed analysis and summary of the Ulysses dust detector data can be found in Krüger et al. [2015], Strub et al. [2015] and Sterken et al. [2015]. During its 8 year long mission in the Jovian system (1995-2003) Galileo accomplished several scientific objectives like e.g., releasing its atmospheric probe or to investigate the Galilean moons. During the mission phase the Dust Detection System analyzed the dust environment in detail [Krüger et al., 2001, 2006, 2010]. IDPs on retrograde orbits and ISD particles have been detected [Colwell et al., 1998] as well as ejecta from the Galilean moons [Krivov et al., 2002]. In its final mission phase the Dust Detection System analyzed also the Gosammer ring before Galileo fell into Jupiter [Krüger et al., 2009].

Soja et al. [2015a] re-analyzed the Dust Detection System data-set in order to determine the dynamical properties of the impacting dust particles. Since the angular FOV of the Dust Detection System was approximately 140° for class 2 Jovian dust stream particles, smaller for

class 3 AR1 Jovian dust stream particles and up to 180° for non-Jovian dust particles [see Krüger et al., 2006, sec. 3], a theoretical impact direction computation is rather difficult. Table 5.2 shows the computational steps how Soja et al. [2015a] computed the orbital elements of the Dust Detection System registered dust particles.

Table 5.2.: Computational procedure to determine orbital elements with the Dust Detection System data. From Soja et al. [2015a].

1. Randomly choose a set of 100,000 orbits from within a population of orbital element limits, each representing a simulated impacting particle.

2. Calculate the true anomaly, and the radial r_r, polar v_Θ, and azimuthal v_ϕ velocity components for each model orbit. For a given (a, e, i) there are four possible ways for a particle to meet Galileo corresponding to dust moving inwards ($r_r < 0$) or out ward ($r_r > 0$) and upward ($v_\Theta < 0$) or downward ($v_\Theta > 0$).

3. We remove orbits that cannot intersect at the current Galileo location. For the remaining orbits, we calculate the jovicentric, Cartesian velocity coordinates v_x, v_y and v_z, and the impact velocities relative to Galileo. At this stage, we have the full set of orbits that can strike the Galileo spacecraft, as well as the relative velocity vector for each possible impact.

4. Calculate the impact angle between the Galileo pointing direction, and the relative velocity vector, and use this to calculate the sensitive area for each of the four orbital solutions using Eq. (2)[3]. Particles that fall within the opening angle of the detector (including correction for wall impacts) are defined to have non-zero sensitive area.

The DA of the CDA has a smaller angular FOV than the Dust Detection System. The angular aperture for the IID is approximately 90° and for the CAT approximately 60°, however, with a decreasing sensitive area for larger impact angles. Figure 5.10 shows the sensitive area (in m^2) for both targets vs. the impact angle in degrees. Due to its smaller angular FOV and with robust statistical measures, pure statistical orbit computations of CDA measurements are sufficient for proper dynamical analysis.

[3]This equation refers to eq. 2 in Soja et al. [2015a] and describes the effective area A_{eff} of the Dust Detection System sensor. It is defined as: $A_{eff}(V_{imp}, \phi, t) = |V_{imp}| / |V_d| \cdot A(\phi, t)$. V_{imp} is the impact velocity, V_d is the jovian-centric velocity of the dust particle. A is the sensitive area vs. the rotation angle ϕ of the spacecraft and the time t. For more geometry details [see Altobelli et al., 2005, sec. 2.2]

5.2.3. Requirements

To compute dynamical properties of in-situ dust particles properly, a few computational requirements need to be determined. Determining the orbit of a dust particle requires 6 measures, due to the 6 degrees of freedom, namely the position and velocity vector of the particle. Adding the detection time leads to the mean anomaly and corresponding epoch. This is listed as requirement **R 2.3** and is stored for all scientific and Housekeeping data.

The position of a detected particle, at the measurement time, is simply determined by the position of the Cassini spacecraft at that particular time (**R 1.1**). These data are available in the SPICE kernel repository and can be transformed to any defined reference frame like the Saturn-centric non-rotating coordinate system (SCNRC). The impact velocity vector $\vec{v}$ of the particle depends on several factors. Equation 5.1 shows that the velocity vector can be separated into two parts, namely the velocity of the particle $\|v_{comb}\|$ and the flight direction $\overrightarrow{n_{fd}}$. Since Cassini's velocity within the Saturnian System is a few km/s, $\|v_{comb}\|$ results from the vectorial addition of Cassini's velocity vector (**R 1.2**), and the flight direction (**R 2.1**) and pace of the impacting particle as seen from CDA (**R 2.2b**).

Cassini's velocity information are also stored in the according kernels. Since these kernels do not provide any error estimate on the provided values, these parameters are assumed to be exact. Also requirement **R 2.1** is provided by SPICE. Boresight information can be obtained from the corresponding C-Kernels. With the articulation angle from the CDA Housekeeping database, the pointing direction of the instrument can be determined in any reference frame.

$$\overrightarrow{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \|v_{comb}\| \cdot \overrightarrow{n_{fd}} \tag{5.1}$$

However, CDA has a certain opening angle and the directional sensitivity depends on the hit target (IID or CAT). Thus, the actual impacting direction cannot be determined precisely, only within a certain probability cone that needs to be defined. This is listed as requirement **R 2.1b** and will be determined and explained in subsection 5.2.4.

Furthermore, the velocity determination is also inaccurate and error-prone. Requirement **R 2.2b** lists the necessity to determine a formulation for the velocity determination error. This is studied in section 5.2.5.

Summarized it can be concluded that most requirements can be fulfilled with the SPICE routines and according kernel repository. Only CDA specific properties need to be determined. Afterwards all requirements are set to determine astro-dynamical properties of in-situ measured dust particles.

R 1.1 Cassini's position

R 1.2 Cassini's velocity

R 2.1 CDA's boresight in SCNRC

R 2.1b PDF of the impacting direction

R 2.2 CDA's determined velocity value of the impacting particle

R 2.2b Error of the determined velocity

R 2.3 Detection time

5.2.4. Impact Direction

In this section the impact directionality of in-situ dust particle measurements is analysed. The resulting CDA specific directional PDFs fulfill requirement **R. 2.1b**.

CDA sensitive area To compute an impact direction probability, the sensitive sensor area A_{sens} of CDA vs. the impact angle needs to be computed. Srama [2000] has computed in detail A_{sens} for the IID and CAT (for more details [see Srama, 2000, sec. 2.2.1]). For this study and the future DOCC Srama [priv. comm.] re-computed the sensitive area of both targets in a high resolution of 0.25°. The generated data set is shown in the appendix C.1. For the study only IID and CAT impacts are taken into account. Possible wall impacts need further investigations. Figure 5.10 shows A_{sens} in m^2 vs the impact direction.

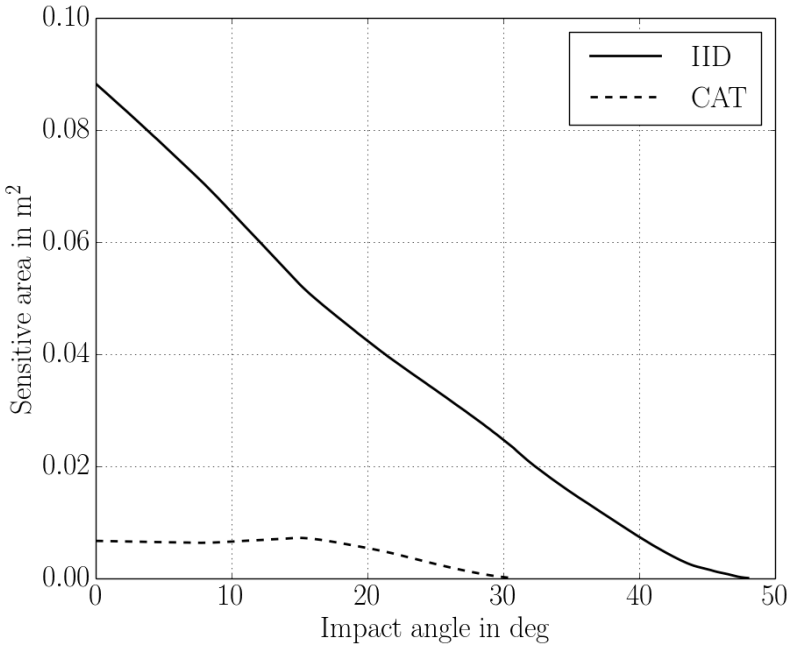


Figure 5.10.: Sensitive area of the Cosmic-Dust-Analyzer. Here, the sensitive areas are shown for the Impact Ionisation Detector (solid line) and the Chemical Analyzer Target (dashed line). The area is given in m^2. Input data can be seen in appendix C.1

0° refers to an impact direction anti-parallel to the boresight. The maximum of the IID function is at 0°. Afterwards, the function decreases almost linearly down to approximately

50°. The sensitive area of CAT is approximately constant between 0° and 15° and decreases afterwards to 0 m^2 at 30°. This is due to the shadowing of the multiplier and its mounting above the CAT. The integrated sensitive area leads to the effective area A_{eff}.

PDF of the impact direction Based on figure 5.10 the PDF of the impact direction of a detected particle can be derived. The normalized sensitivity area of IID and CAT vs. the impact angle can be interpreted as an impact probability of a dust particle. The median of the probability density function is 12° (for both targets), thus, half of the detections come within a full cone angle of 24°. To derive an impact direction probability, first, the function of figure 5.10 needs to be rotated by 2π, due to the cylindric shape of the detector's tube. The resulting function is then normalized. Figure 5.11 and 5.12 show the PDF in an angle-angle space. The radially shown angle refers to the impact angle shown in figure 5.10. The angular 2π space results from the axial symmetric CDA shape. The star-shaped mounting of the multiplier (which can be seen in figure 5.1 b)) is already computed for the sensitive area. Thus, the shown PDFs are slightly simplified.

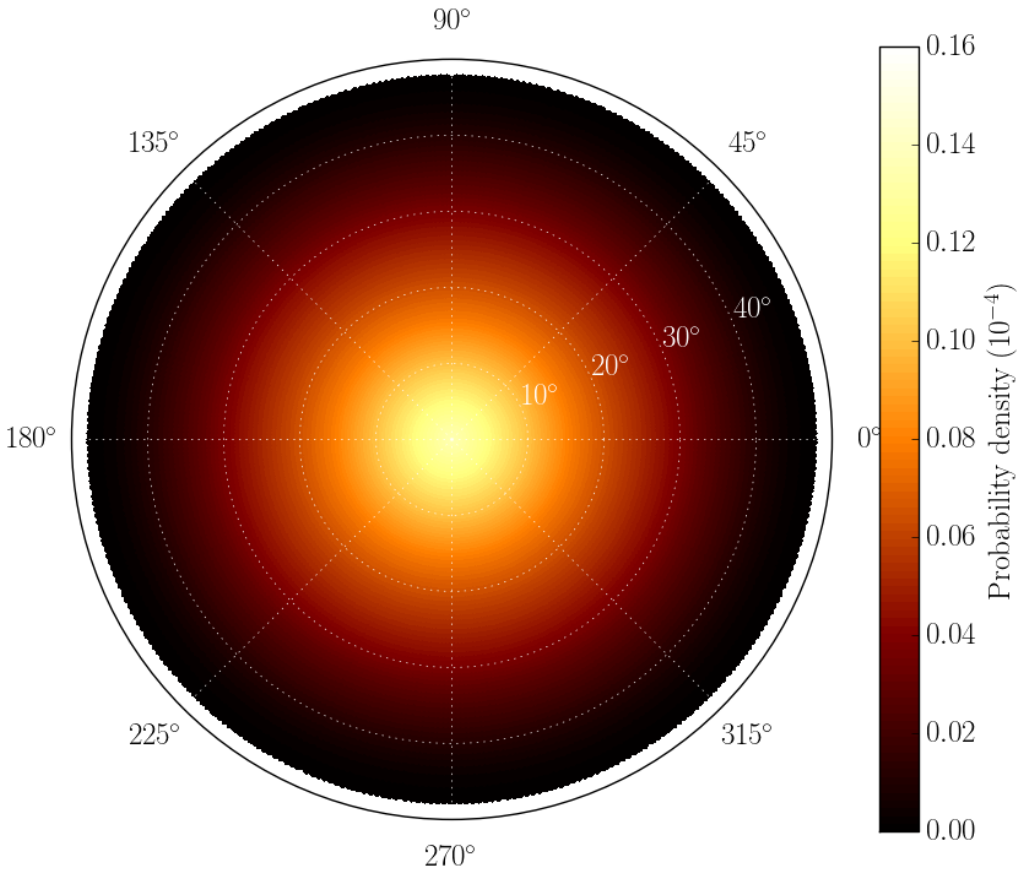


Figure 5.11.: Probability Density Function of the impact direction of a dust particle on the Impact Ionisation Detector. The radially shown angle is the impact angle between particle direction and the anti-parallel boresight vector. Since the Cosmic-Dust-Analyzer has a cylindric shape, the impact probability is axially symmetric (0° to 360°). The color code shows the Probability Density Function. Integrated over the entire space, the probability is 1.

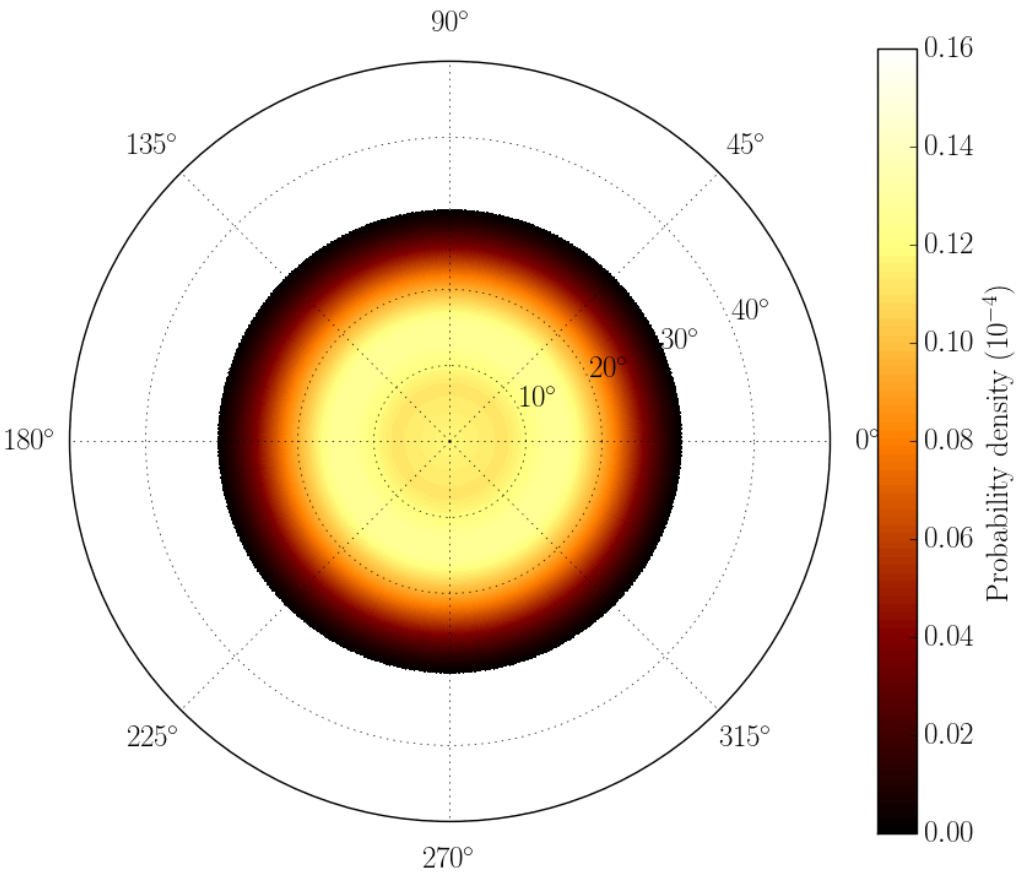


Figure 5.12.: Probability Density Function of the impact direction of a dust particle on the Chemical Analyzer Target. The radially shown angle is the impact angle between particle direction and the anti-parallel boresight vector. Since the Cosmic-Dust-Analyzer has a cylindric shape, the impact probability is axially symmetric (0° to 360°). The color code shows the Probability Density Function. Integrated over the entire space, the probability is 1.

Computational procedure A detected dust particle has a certain impact direction probability. Since the corresponding PDF is derived numerically, and the link between the PDF of the impact direction and resulting orbital elements cannot be described in a simple analytical form, the DOCC needs a MC algorithm. Based on the boresight of CDA at the detection time of an impact event, a probability cone around the boresight vector is needed to compute additional artificial particles at this time stamp. Thousands or Hundred-thousands re-sampled trajectories need to be computed to determine the statistics of the particle dynamics. The probability P is a function of two angles δ and α, where δ describes the angle between anti-parallel boresight and impact direction and α ranges within the axially symmetric 2π space. We have the two PDFs $P_{iid}(\delta, \alpha)$ and $P_{cat}(\delta, \alpha)$. For both functions we have:

$$\int_{0°}^{360°} \int_{0°}^{48.25°} P_{iid}(\delta, \alpha) \; \mathrm{d}\delta \mathrm{d}\alpha = 1 \tag{5.2}$$

$$\int_{0°}^{360°} \int_{0°}^{31.00°} P_{cat}(\delta, \alpha) \; \mathrm{d}\delta \mathrm{d}\alpha = 1 \tag{5.3}$$

$P_{iid}(\delta, \alpha)$ and $P_{cat}(\delta, \alpha)$ are joint probability distributions, since the original function $P_{iid}(\delta)$ and $P_{cat}(\delta)$ are rotated by 2π. We have:

$$P_{iid}(\delta, \alpha) = P_{iid}(\delta) P_{iid}(\alpha) \tag{5.4}$$

$$P_{cat}(\delta, \alpha) = P_{cat}(\delta) P_{cat}(\alpha) \tag{5.5}$$

$$\tag{5.6}$$

Thus, for the computation procedure, the probability depending on δ and α can be computed independently. For $P_{iid}(\delta)$ and $P_{cat}(\delta)$, the normalized distributions shown in figure 5.10 are used. Since both computations are actually discrete functions with a resolution of 0.25°, the resulting PDFs become also discrete. Figure 5.13 and 5.14 show an ideal re-sampling case based on $P_{iid}(\delta)$ and $P_{cat}(\delta)$. The re-sampling shows the resulting distributions for IID and CAT, respectively. Each impact angle bin corresponds to the resolution size of 0.25°. In both figures, the last bin (IID: 48.00° to 48.25°; CAT: 30.75° to 31.00°) contains only one sample. The white solid line shows the median of the impact angle and the white dashed lines show the lower and upper limit of the Inter Quartile Range (IQR). The median shows the 50 % separation of a function, the lower and upper limit of the IQR show the 25 % and 75 % separation, respectively.

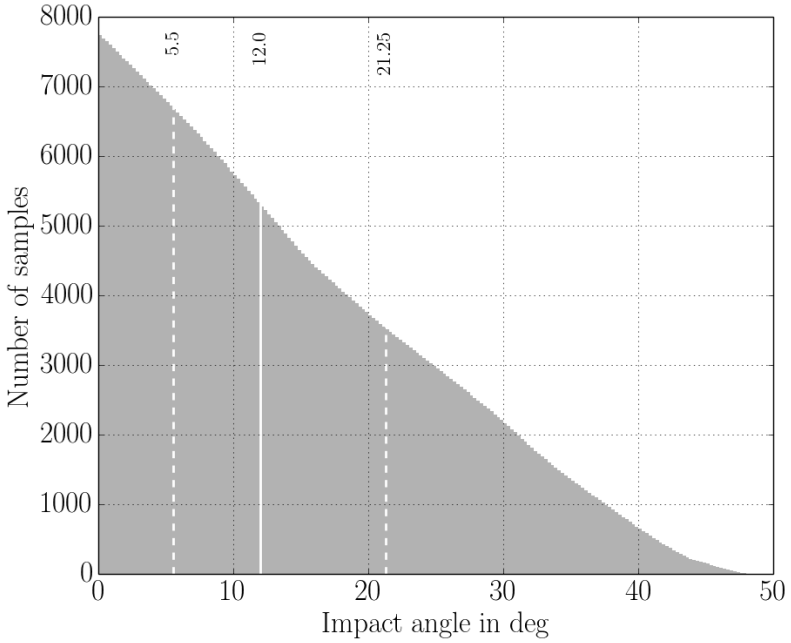


Figure 5.13.: Ideal re-sampling of the Probability Density Function $P_{iid}(\delta)$. Here, the re-sampling parameter have been artificially set to get 1 re-sampling in the last bin ranging from 48.00° to 48.25°. The solid line indicates the median and the dashed lines show the corresponding Inter Quartile Range limits. The values above the lines indicate the corresponding angles. The total number of re-samples is: 639,819.

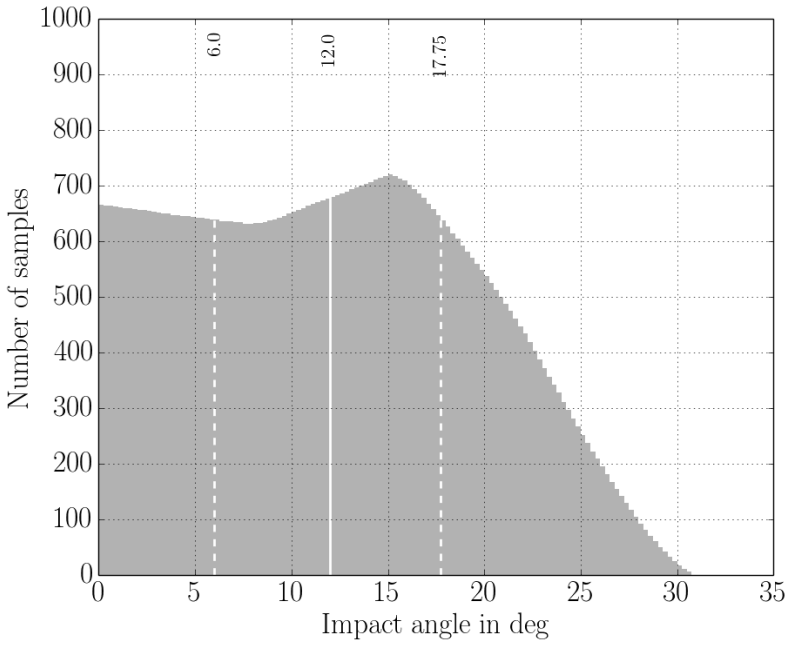


Figure 5.14.: Ideal re-sampling of the Probability Density Function $P_{cat}(\delta)$. Here, the re-sampling parameter have been artificially set to get 1 re-sampling in the last bin ranging from 30.75° to 31.00°. The solid line indicates the median and the dashed lines show the corresponding Inter Quartile Range limits. The values above the lines indicate the corresponding angles. The total number of re-samples is: 63,150.

For the PDF $P_{iid}(\alpha)$ and $P_{cat}(\alpha)$ the PDFs is a discrete uniform distribution with $\alpha \in (0, 360]$. Additionally, the step size of α corresponds to the resolution of the PDFs depending on δ, namely 0.25°.

5.2.5. Velocity Determination

This section analyses the requirements **R. 2.2** and **R. 2.2b**.

Measured Velocity Srama [2009] described in section 3.4.1 and 3.5.1 the velocity determination for the IID and CAT, respectively. For the study the rise time t_r is used from the QI and QC channel. Equation 3.14 and 3.29 in Srama [2009] show the conversion functions to determine the velocity of the impacting particle and is shown here in equation 5.7 and 5.8. The rise time from the QI channel is used for IID impacts and the rise time from the QC channel is used for the CAT impacts. t_r is given in microseconds and the resulting velocity is in km/s.

$$v_{IID}^{QI}(t_r) = -55.3 + 253 \cdot t_r^{-0.423} + 0.222 \cdot t_r \tag{5.7}$$

$$v_{CAT}^{QC}(t_r) = 1.7 + 19.2 \cdot t_r^{-0.979} - 0.0204 \cdot t_r \tag{5.8}$$

Both functions have been derived from calibration measurements at the Dust Acceleration Facility at the Max Planck Institute for Nuclear Physics. A description of the set-up can be

read in Srama [2009]. Figure 5.15 and 5.16 show the calibration data for IID and CAT impacts, respectively. The velocity has been derived from the Particle Selection Unit (PSU) and the rise times have been determined from the QI and QC channels. The total number of measurements was 712 for IID and 197 for CAT (actually 198, but one erroneous $t_r = 0.0$ μs measurement is within the data set). Equation 5.7 is valid for rise times between 1.0 μs and 75.0 μs and particle charges larger than 20 fC [Srama, 2009]. Equation 5.8 is valid for particles charges larger than 10 fC [Srama, 2009]. A rise time limit has not been set by Srama [2009], however, based on the calibration data range the valid range appears to be within 0.2 μs and 30 μs.

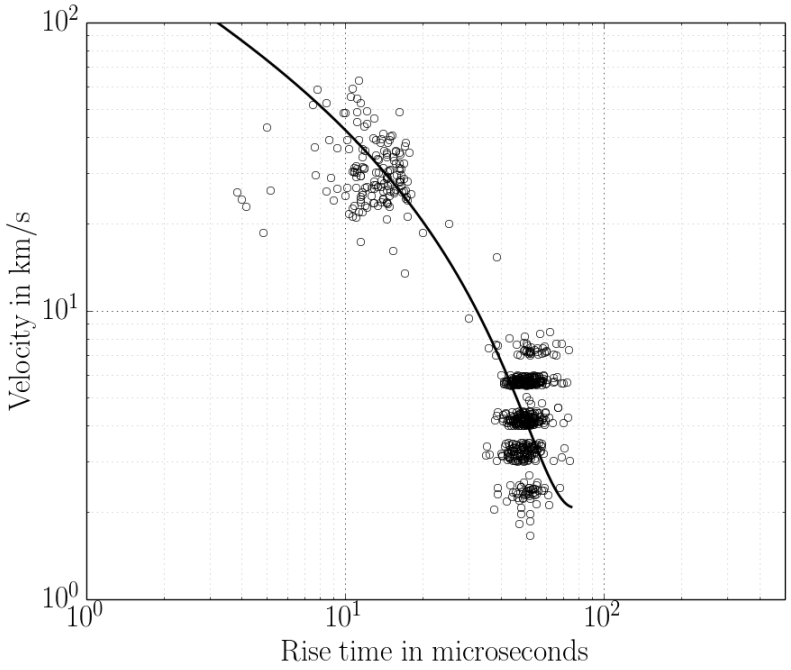


Figure 5.15.: Velocity vs. rise time of the QI channel for dust impacts on the Impact Ionisation Detector. The dots show the velocity calibration data. The velocity determination has been done by the Particle Selection Unit. The fit curve is shown in equation 5.7 and is valid for rise times between 1.0 μs and 75.0 μs and for particles with a charge of at least 20 fC.

Thus, for the study and the DOCC equation 5.7 and 5.8 can be used to determine the impact velocity of a particle. Using both equations, one needs to take the rise time and charge of the particle into account.

The first version of the CDA science database [Albin, 2014] did not include the parameters of the particles which are needed to compute the velocity. Thus, table *cda_events* has been updated. The following list shows the new entries in the database. The entries correspond to table entry number 58 - 66.

58 *cda_qc_tr*: Rise time QC channel signal in seconds

59 *cda_qt_tr*: Rise time QT channel signal in seconds

60 *cda_qi_tr*: Rise time QI channel signal in seconds

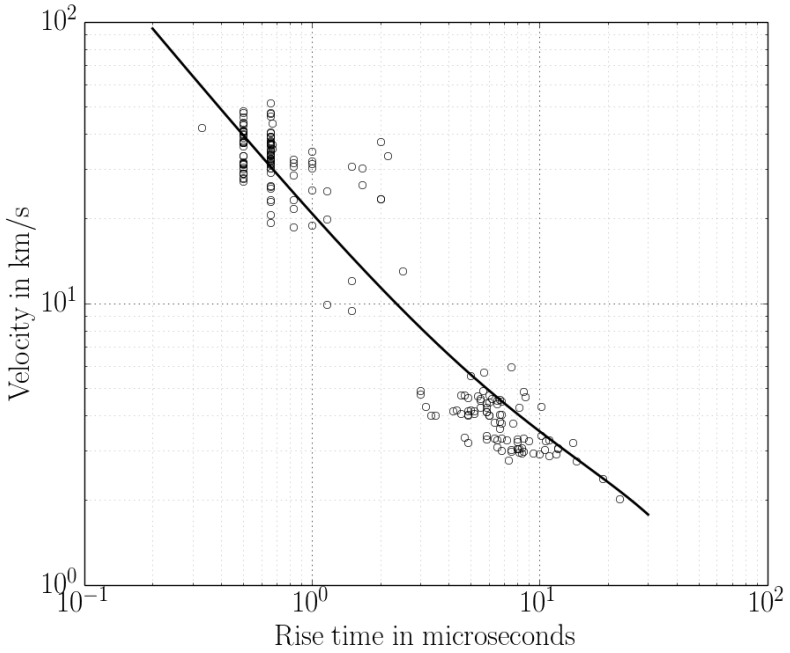


Figure 5.16.: Velocity vs. rise time of the QC channel for dust impacts on the Chemical Analyzer Target. The dots show the velocity calibration data. The velocity determination has been done by the Particle Selection Unit. The fit curve is shown in equation 5.8 and is valid for rise times between 0.2 μs and 30.0 μs and for particles with a charge of at least 10 fC.

61 *cda_qc_amp*: Amplitude QC channel signal in Coulomb

62 *cda_qt_amp*: Amplitude QT channel signal in Coulomb

63 *cda_qi_amp*: Amplitude QI channel signal in Coulomb

64 *cda_qc_int*: Integral QC channel signal

65 *cda_qt_int*: Integral QT channel signal

66 *cda_qi_int*: Integral QI channel signal

Thus, with the database table entries 58 - 66 in table *cda_events* and equation 5.7 and 5.8, requirement **R 2.2** is fulfilled.

Velocity Error Figure 5.15 and 5.16 show that the determined velocities scatter around the mean fit. Thus, for the velocity determination an error needs to be computed, see requirement **R 2.2b**. Here, a similar method is used as described and shown in figure 3.19 and 3.28 in Srama [2009]. Srama [2009] determined the error by comparing the theoretically expected velocities with the measured velocity values.

The developed procedure is explained in the following. Figure 5.17 and 5.18 show the normalized distribution of the theoretical velocity data divided by the measured velocities ($\tilde{v} = v_{theory}/v_{PSU}$) for IID and CAT, respectively. The histogram bars show the distribution with a resolution of 0.1. The solid lines are the computed PDFs $P_{v,iid}$ and $P_{v,cat}$ for IID and

CAT, respectively. These functions are derived with a Kernel Density Estimator (KDE) and the following bandwidth given by Scott [2015]:

$$b_{gauss} = 1.144 \cdot \sigma \cdot n^{-1/5} \tag{5.9}$$

b is the bandwidth, σ is the standard deviation of the data set and n is the number of data points. Using a *Gaussian* kernel, which is defined from minus to plus infinity, would lead to:

$$\int_{-\infty}^{0} P_{v,iid;cat} \geq 0 \tag{5.10}$$

For the error computation, this would lead to negative velocities. Thus, the *Epanechnikov* kernel is chosen, whose domain is restricted. Table 6.3 in Scott [2015] provides a list which provides coefficients to convert the bandwidth determined by equation 5.9 to bandwidths for different kernels. Equation 5.11 shows the "rule-of-thumb" bandwidth for the *Epanechnikov* kernel.

$$b_{epan} = 2.214 \cdot b_{gauss} \tag{5.11}$$

The resulting KDE fit is shown as the black solid line in both figures 5.17 and 5.18. The white solid and dashed lines show the median and corresponding IQR, respectively. Median, the lower and upper boundary of the IQR are the 50^{th} (separating the data in half), 25^{th} (separating the data in a lower 25 % and higher 75 % set) and 75^{th} (separating the data in a lower 75 % and higher 25 % set) percentile of a data set. Thus, the IQR contains 50 % of the data around the median. A perfect Gaussian distribution would lead to symmetric IQR limits. Here, the median and IQR for IID and CAT are $\tilde{v}_{IID,med} = 1.01^{+0.31}_{-0.26}$ and $\tilde{v}_{CAT,med} = 1.08^{+0.22}_{-0.20}$, respectively.

Similarly as described in the section before, the velocity error PDF is used as an input function to generate several dust particle velocities. E.g., with the directional PDF 10,000 particle directions are generated. Then, based on the signal rise time, the mean velocity is computed. The velocity is then used to scale the velocity error PDF. Based on the generated function, 10,000 velocities are re-sampled and linked with the flight direction vectors.

To improve the statistics of figure 5.17 and 5.18 an outlier determination algorithm is applied on the data. To detect outliers or extreme values several methods can be used: supervised or un-supervised methods, and methods based on a statistical approach or by modeling. Here, a simple approach is used by Tukey [1977], where he defines data points which lay 1.5 times the IQR above the upper or below the lower IQR limit, as outliers or extreme values. Figure 5.19 shows the box plot of the $\tilde{v}_{IID}$ and $\tilde{v}_{CAT}$ data. The box indicates the IQR with the horizontal

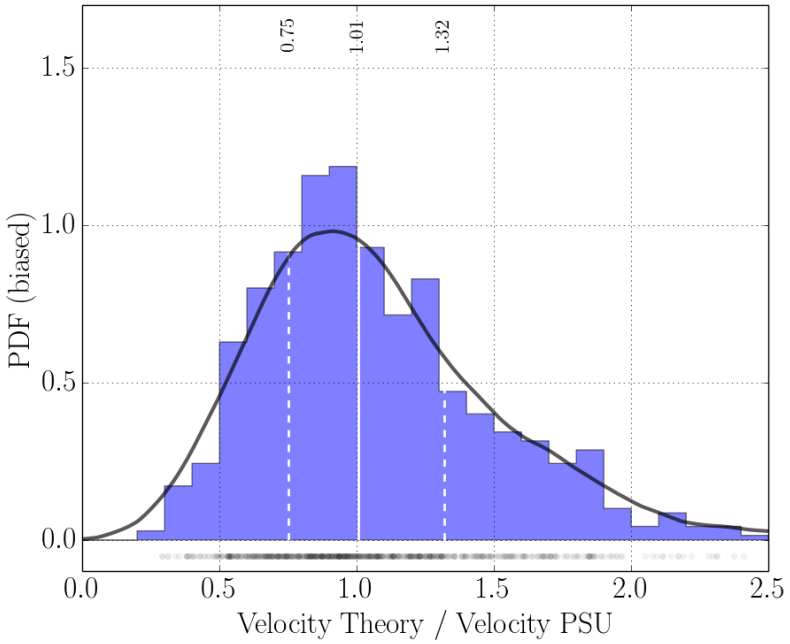


Figure 5.17.: Velocity determination error $\tilde{v} = v_{theory}/v_{PSU}$ distribution of the Impact Ionisation Detector. The histogram bars have a resolution of 0.1, the black solid line is a Kernel Density Estimator, based on an *Epanechnikov* kernel. The black dots show the actual data points and the white solid line and the dashed lines indicate the median and corresponding Inter Quartile Range ($1.01^{+0.31}_{-0.26}$). Data from Srama [2009].

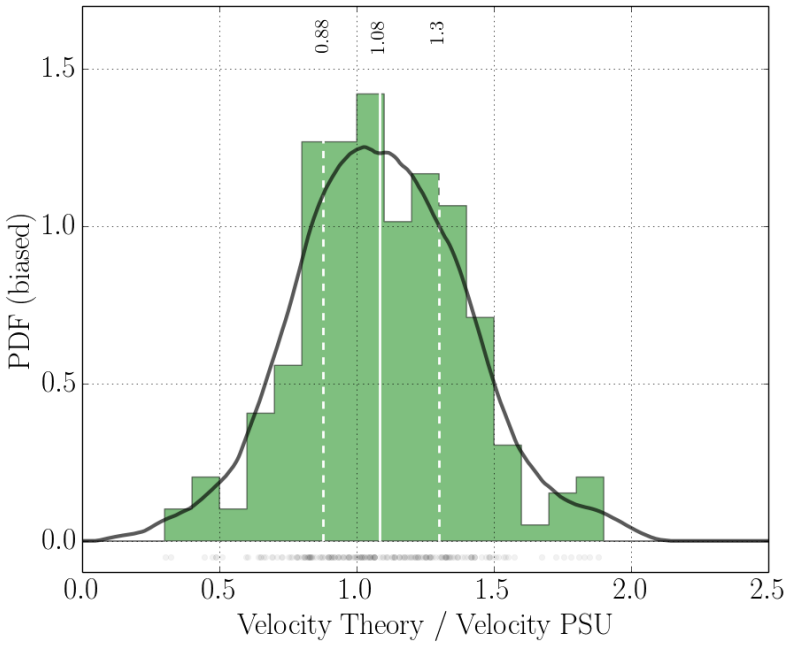


Figure 5.18.: Velocity determination error $\tilde{v} = v_{theory}/v_{PSU}$ distribution of the Chemical Analyzer Target. The histogram bars have a resolution of 0.1, the black solid line is a Kernel Density Estimator, based on an *Epanechnikov* kernel. The black dots show the actual data points and the white solid line and the dashed lines indicate the median and corresponding Inter Quartile Range ($1.08^{+0.22}_{-0.20}$). Data from Srama [2009].

line as the median. The dashed lines are the 1.5 IQR range above and below the IQR limits. The dots show the outliers. On the very right of each box plot is the number of outliers shown.

Based on this criterion CAT and IID contain 2 and 25 outliers, respectively.

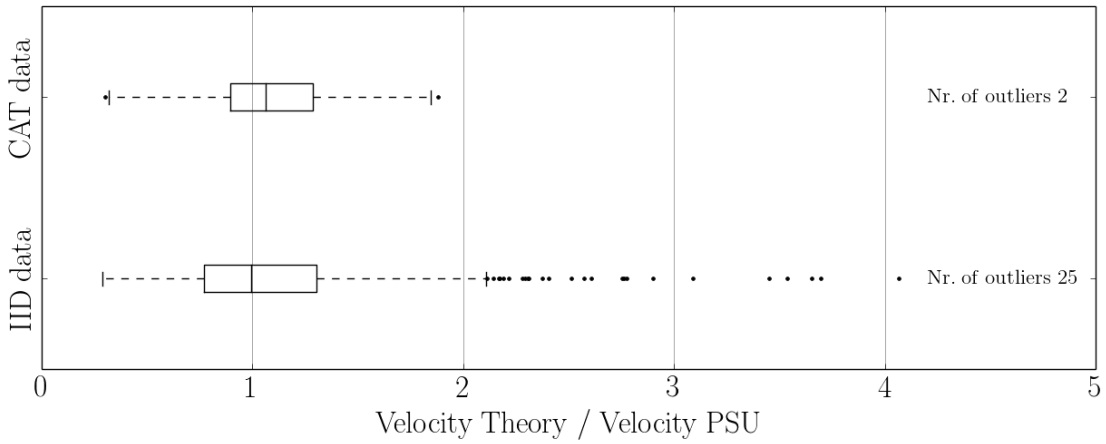


Figure 5.19.: Box plots of the error factors for $\tilde{v}_{IID}$ and $\tilde{v}_{CAT}$ data. The box and horizontal line indicate the Inter Quartile Range and median, respectively. The dashed lines show the 1.5 times the Inter Quartile Range range above and below the Inter Quartile Range limits and the dots show the outliers. 2 and 25 outliers are within the Chemical Analyzer Target and Impact Ionisation Detector data sets, respectively.

Based on the outlier criterion by Tukey [1977], figure 5.17 and 5.18 are re-analyzed and are shown in figure 5.20 and 5.21, respectively. Since only 2 CAT values are sort out, the resulting un-biased CAT PDF remains the same. The median and IQR are $\tilde{v}_{CAT,med} = 1.08^{+0.22}_{-0.20}$. Data from the IID measurements are spread up to $\tilde{v}_{IID} > 2.0$. Most identified extreme values lay in the wide spread part of the function shown in figure 5.17. The re-analysis shows a more narrowly shaped distribution with a median and IQR of $\tilde{v}_{CAT,med} = 0.99^{+0.30}_{-0.24}$. Both determined velocity PDFs which are shown in the following two figures are used for the MC velocity re-sampling.

However, Srama [2009] derived the velocity equations 5.7 and 5.8 with the complete data set. For CAT only 2 data points (1 % of the data) and for IID 25 data points (3.5 % of the data) are neglected. Furthermore, to allow a better comparison with previous results and publications, the velocity functions depending on the rise time are not being re-analyzed.

The unbiased velocity distributions lead to the following Median Absolute Deviation (MAD) values:

$$v_{\mathrm{MAD,IID}} = 0.25 \text{ km/s} \tag{5.12}$$

$$v_{\mathrm{MAD,CAT}} = 0.20 \text{ km/s} \tag{5.13}$$

Assuming a Gaussian distribution, the MAD values need to be multiplied by a factor of 1.4826 to compute the corresponding standard deviation. The velocity PDF of the IID has a skewness towards a higher ratio where the CAT distribution appears more symmetrically. Applying the factor to both MAD values leads to a standard deviation σ of:

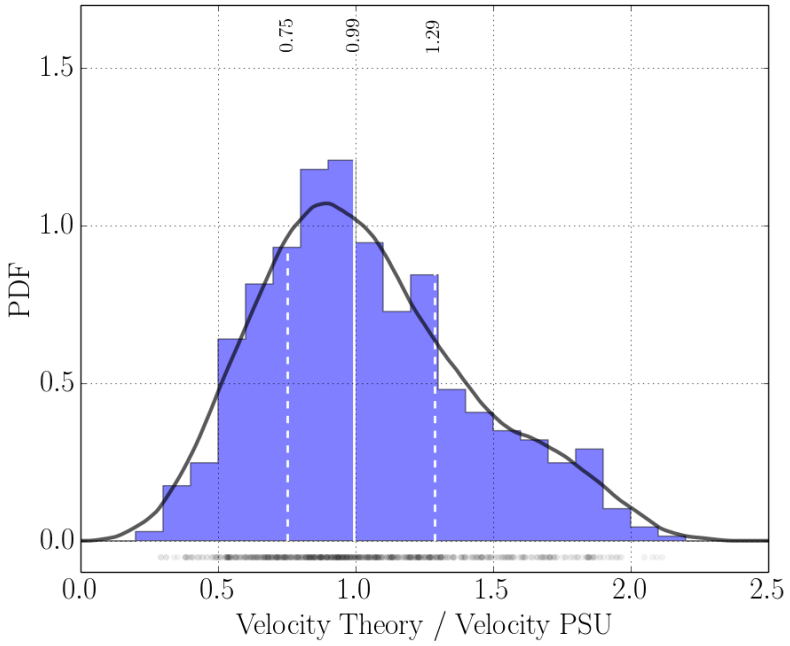


Figure 5.20.: Velocity determination error $\tilde{v} = v_{theory}/v_{PSU}$ distribution (Impact Ionisation Detector, unbiased). The histogram bars have a resolution of 0.1, the black solid line is a Kernel Density Estimator, based on an *Epanechnikov* kernel. The black dots show the actual data points and the white solid line and the dashed lines indicate the median and corresponding Inter Quartile Range ($0.99^{+0.30}_{-0.24}$). Data from Srama [2009].

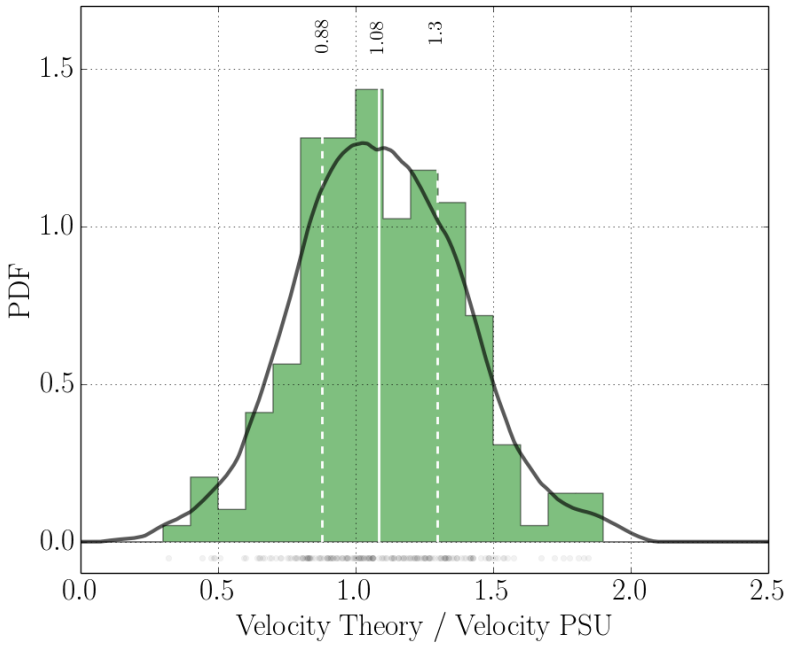


Figure 5.21.: Velocity determination error $\tilde{v} = v_{theory}/v_{PSU}$ distribution (Chemical Analyzer Target, unbiased). The histogram bars have a resolution of 0.1, the black solid line is a Kernel Density Estimator, based on an *Epanechnikov* kernel. The black dots show the actual data points and the white solid line and the dashed lines indicate the median and corresponding Inter Quartile Range ($1.08^{+0.22}_{-0.20}$). Data from Srama [2009].

$$\sigma_{\mathrm{IID}} = 0.37 \text{ km/s} \tag{5.14}$$

$$\sigma_{\mathrm{CAT}} = 0.30 \text{ km/s} \tag{5.15}$$

Thus, the standard deviation is approximately 1/3 of the determined mean value.

Computational procedure As described in the section before, first of all, an ensemble of particle directions for each impact event is computed. Each vector is normalized and does not contain any velocity information. Depending on the hit target and needed channel rise time (IID – QI; CAT – QC) a mean velocity is computed using equation 5.7 and 5.8, respectively. The computed mean velocities are then used to scale the normalized velocity PDFs which are shown in figure 5.20 for the IID and 5.21 for the CAT.

Four examples of these scaled PDFs are shown in figure 5.22 – 5.25. The first two figures show the velocity PDFs for IID with mean velocities ranging from 1 km/s to 5 km/s and the second from 10 km/s to 50 km/s. The last two images show the same distributions for particles impacting on the CAT. The color code corresponds to a measured mean velocity and the according colorbar is shown on the right panel of each figure. The width of the PDF increases for higher impact velocities. Presumably, a high velocity variation affects the statistics of velocity depending orbital elements significantly, like e.g., the eccentricity and consequently the semi-major axis.

Based on these scaled PDFs a routine generates velocity values to scale the normalized direction vectors of the MC particle ensemble.

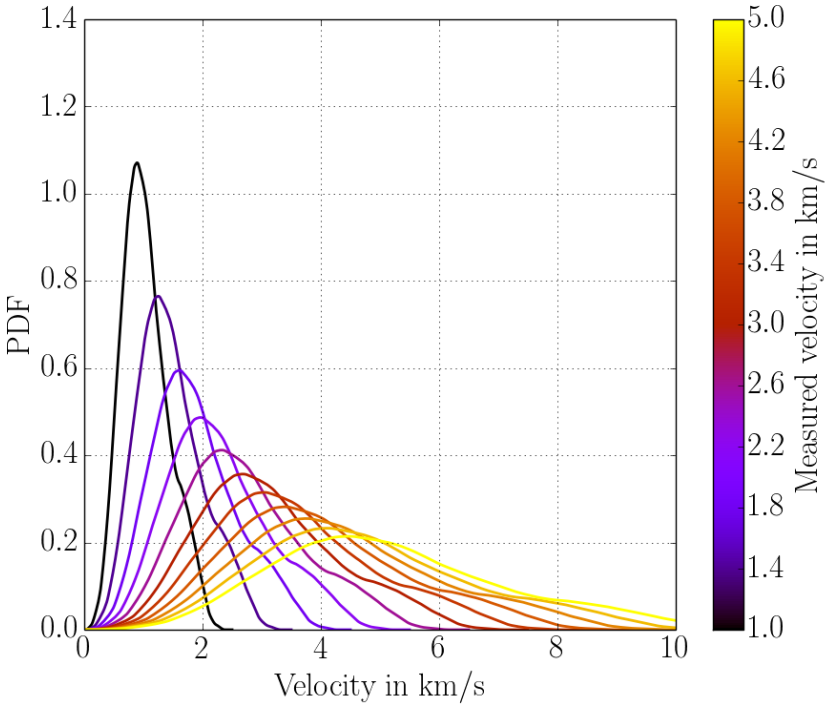


Figure 5.22.: Velocity Probability Density Function for impacts on the Impact Ionisation Detector target. The color code corresponds to the measured mean velocity. E.g., the black curve shows the velocity Probability Density Function for a particle with a rise time corresponding speed of 1 km/s.

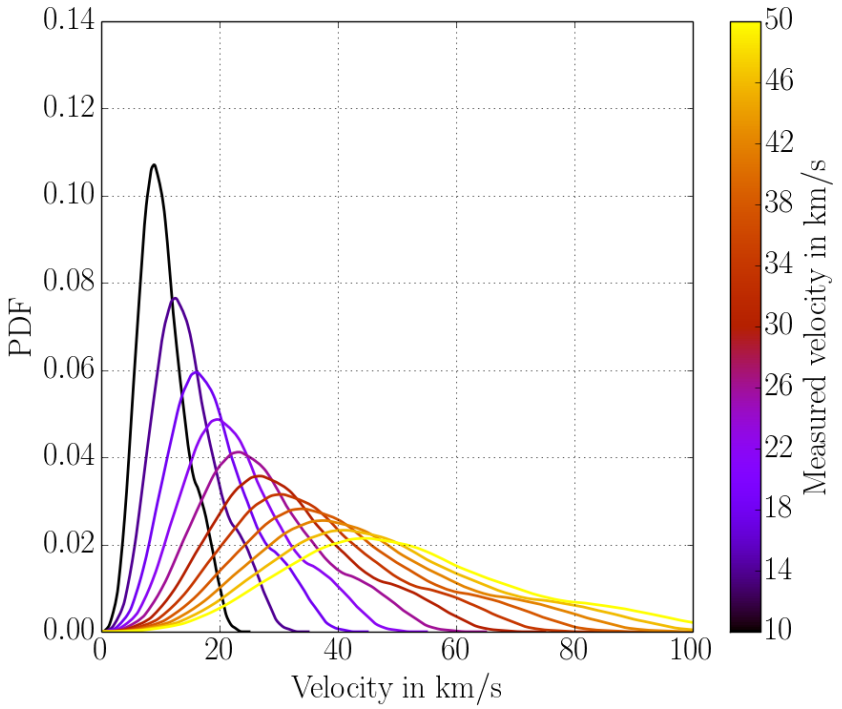


Figure 5.23.: Velocity Probability Density Function for impacts on the Impact Ionisation Detector target. The color code corresponds to the measured mean velocity. E.g., the black curve shows the velocity Probability Density Function for a particle with a rise time corresponding speed of 10 km/s.

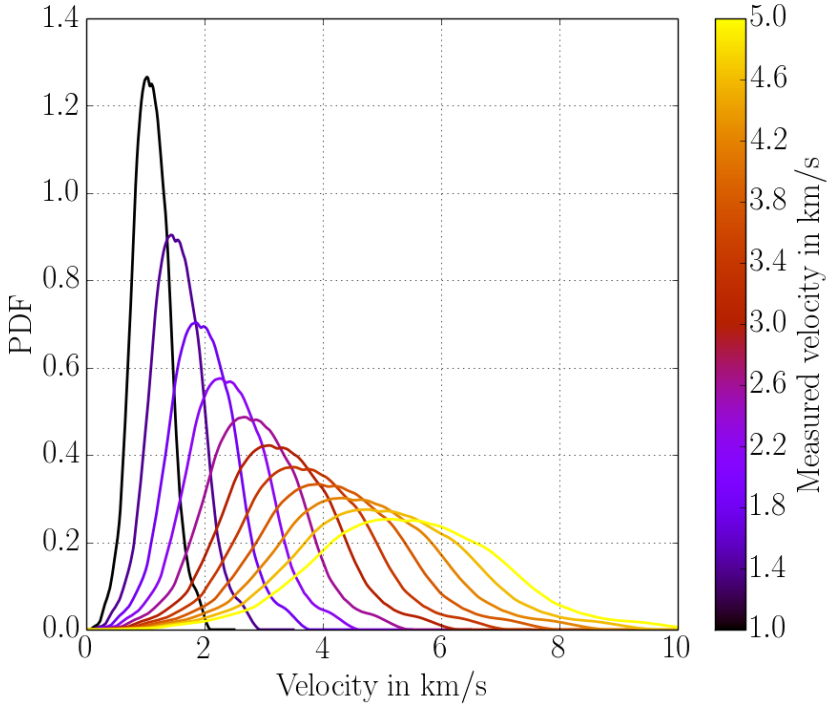


Figure 5.24.: Velocity Probability Density Function for impacts on the Chemical Analyzer Target target. The color code corresponds to the measured mean velocity. E.g., the black curve shows the velocity Probability Density Function for a particle with a rise time corresponding speed of 1 km/s.

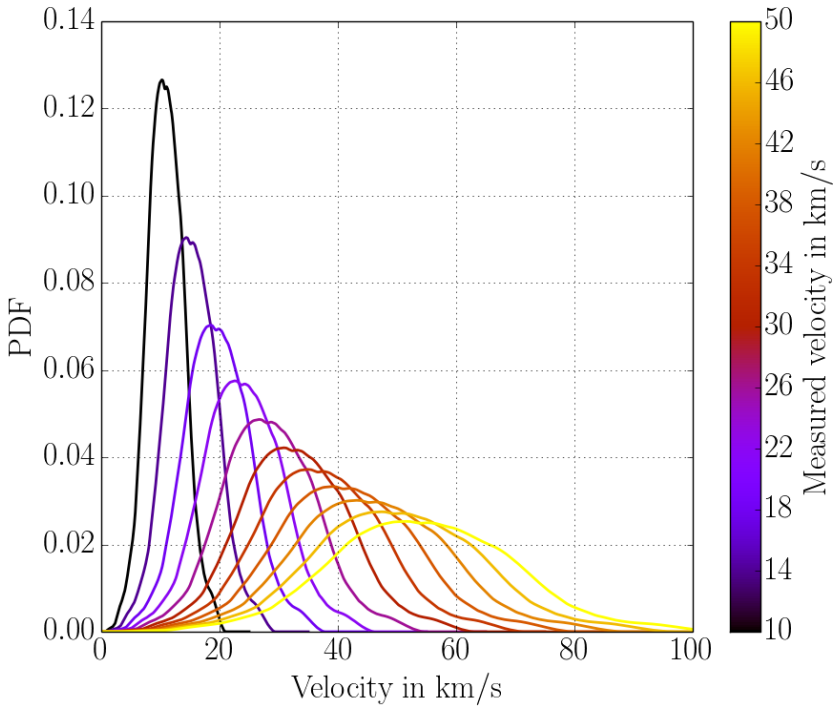


Figure 5.25.: Velocity Probability Density Function for impacts on the Chemical Analyzer Target target. The color code corresponds to the measured mean velocity. E.g., the black curve shows the velocity Probability Density Function for a particle with a rise time corresponding speed of 10 km/s.

5.2.6. Feasibility

With all requirements fulfilled to determine orbital elements of in-situ particles, a few artificial test cases are computed, mainly based on the velocity PDF to analyze the expected solution spaces. This section shows parts of a feasibility study during the development of the DOCC. With all data, kernels and PDFs the section afterwards shows the developed MC based algorithm to determine orbital elements of in-situ dust instrument measurements.

Circular orbits For the shown test cases we set the following conditions, to determine the feasibility of computing circularly orbiting dust particles around Saturn.

- Cassini's position is in Saturn's ring plane
- Cassini is in a "steady state", thus, Cassini does not move within the Saturnian system
- Particles revolve around Saturn on the Ring Plane (no inclination)
- Particles revolve around Saturn on a circular Kepler-orbit
- CDA's boresight is aligned to the Dust-RAM direction
- The mean detection velocity corresponds with the actual Dust-RAM velocity

- All particles impact on CDA anti-parallel to its boresight (no FOV)

- In a later approach, we increase Cassini's velocity with respect to the Dust-RAM direction

The velocity of a circularly orbiting body v_{circ} is constant and is described by equation 5.16, where G is the gravitational constant, m_{Saturn} is the mass of Saturn. Since the mass of Cassini (and surrounding dust particles) is small compared to Saturn, only the mass of Saturn is considered: $\mu_{\text{Saturn}} = G \cdot m_{\text{Saturn}}$. r is the distance between Cassini and the barycenter of Saturn.

$$v_{\text{circ}} = \sqrt{\frac{G \cdot m_{\text{Saturn}}}{r}} = \sqrt{\frac{\mu_{\text{Saturn}}}{r}} \tag{5.16}$$

Figure 5.26 visualizes the Dust-RAM velocity in km/s vs. the radial distance from Saturn in Saturn Radius / Radii (SR). The black curve shows the function 5.16. Assuming that a particle hit the IID or CAT target with the Dust-RAM pace (steady, non-moving observer), the dashed lines show the corresponding IQR measurement range. As described in the section before, the IQR width depends on the impacting velocity. Smaller velocities lead to a smaller detection uncertainty which can be seen in the figure 5.26. The IQR functions converge slightly with the expected Dust-RAM function for larger distances.

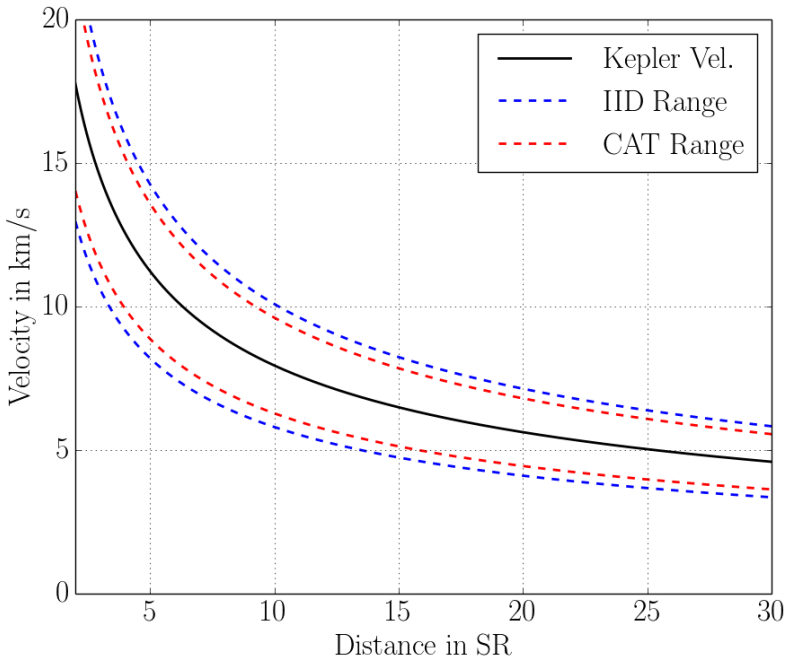


Figure 5.26.: Dust-RAM velocity determination range. The black curve shows the Dust-RAM velocity in km/s vs. the radial distance of a particle from Saturn in Saturn Radius / Radii (SR). The corresponding function is shown in equation 5.16. The red and blue dashed lines show the corresponding IQR measurement range for particles impacting the Impact Ionisation Detector (IID) and the Chemical Analyzer Target (CAT) with the Dust-RAM velocity, respectively (steady, non-moving observer).

Results Since these simple simulations do not include the FOV of both CDA targets and the boresight is aligned parallel to the ring plane, the inclination is 0°. Argument of periapsis and the longitude of ascending node are not computed, since these parameters cannot be determined for circular moving particles with no inclination. Here, the velocity-sensitive orbital elements are shown namely the eccentricity and the semi-major axis.

Figure 5.27 to 5.30 show the results for the semi-major axis (in SR) and eccentricity for CAT and IID impacts, respectively. These plots show only the solution of a steady-state case, where Cassini's velocity is a zero vector in the SCNRC system. The solid and dashed lines show the median and corresponding IQR boundaries. Since the resulting distribution from the MC simulations do not follow a Gaussian function, the Median and IQR are the preferred descriptive statistical parameters to visualize the results.

The generated particles revolve around Saturn in a perfect circular orbit. Thus, theoretically the determined eccentricity should be 0 and the semi-major axis should correspond to the distance from Saturn. The determined semi-major axis median follows approximately the distance of Cassini from Saturn (CAT and IID). However, the IQR of the computation increases for larger distances although the velocity error decreases. This can be explained by figure 5.26. The Dust-RAM velocity function slope decreases for larger distances from Saturn since the velocity is proportional to $1/\sqrt{r}$, where r is the distance between the spacecraft and Saturn's barycenter. Considering the velocity determination error (dashed blue and red lines in figure 5.26) it can be seen that for larger distances the velocity variation corresponds to a larger distance and thus semi-major axis range.

MC results of the eccentricity show a very large variation of the values for all distances. The Median of the eccentricity is 0.4 and 0.5 for CAT and IID, respectively. Considering the IQR, apparently only a small fraction of all computations determined a circular orbit and eccentricity of 0, respectively. Due to the skewness of the IID's velocity PDF the median line of the semi-major axis has a smaller slope than theoretically expected.

The results of the steady-state case show that the Median of the semi-major axis corresponds well with the prediction. Considering the eccentricity results and the large IQR boundaries, the steady-state case however shows that the orbital elements vary too much to determine a proper particle orbit. Consequently, flight dynamical properties, the determination of possible source regions and parent bodies etc. have a too large uncertainty for a proper scientific analysis.

The shown case however is not suitable to estimate the quality of a MC orbit computation code, since the shown geometry is too simplified. Cassini revolves Saturn in eccentric orbits, thus with a certain velocity in the SCNRC. The following four plots show the MC results as shown in figure 5.27 to 5.30 but with different spacecraft velocities with respect to the Dust-RAM velocity. Each sub-figure a) to d) shows the determined orbital elements where the spacecraft has a velocity of 0.25, 0.50, 0.75 and 0.90 with respect to the Dust-RAM velocity in Dust-RAM direction. Thus, particles from the Dust-RAM direction impact with a slower velocity namely with a factor of 0.75, 0.50, 0.25 and 0.10 with respect to the Dust-RAM velocity in

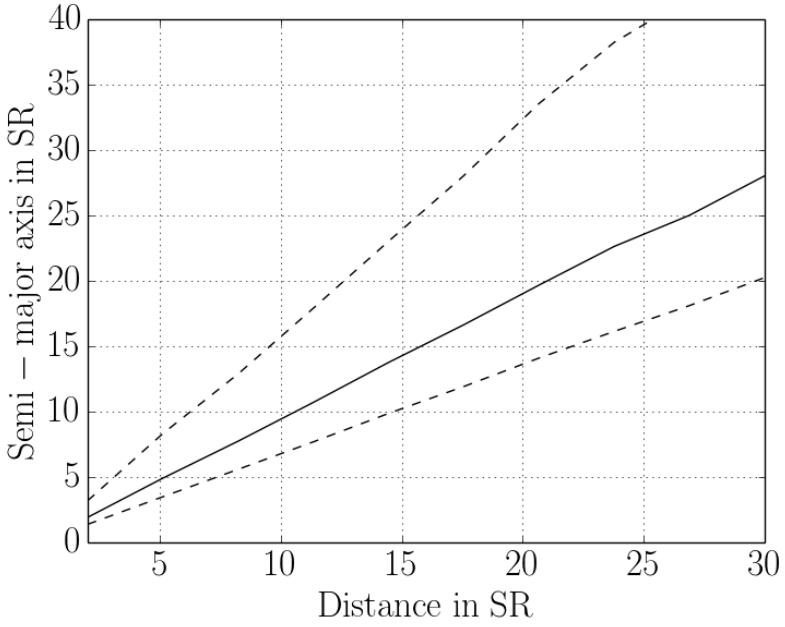


Figure 5.27.: Result of the Monte-Carlo simulation to determine the semi-major axis (Chemical Analyzer Target). The solid and dashed lines show the Median and the Inter Quartile Range of the results vs. the distance from Saturn in Saturn Radius / Radii (SR).

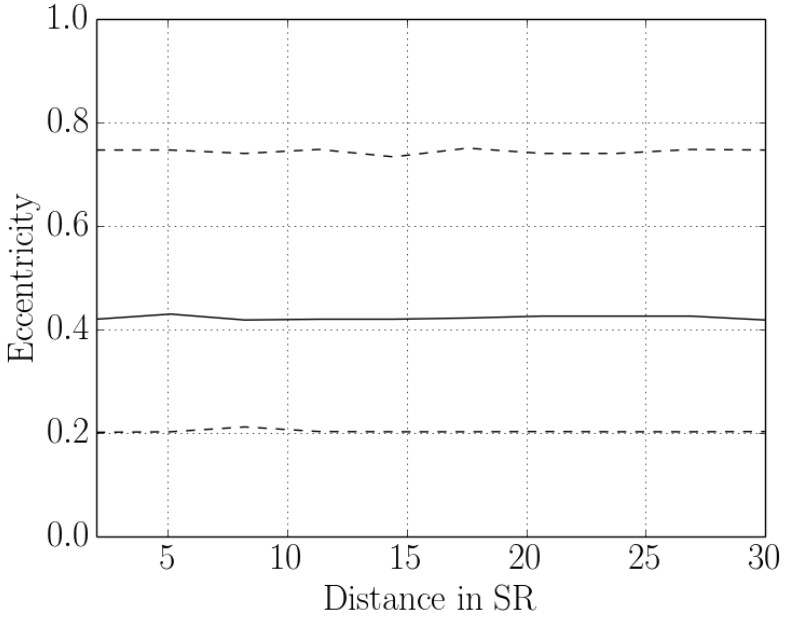


Figure 5.28.: Result of the Monte-Carlo simulation to determine the eccentricity (Chemical Analyzer Target). The solid and dashed lines show the Median and the Inter Quartile Range of the results vs. the distance from Saturn in Saturn Radius / Radii (SR).

SCNRC. To compute the orbital elements, the speed and corresponding error of the impacting particle is computed (with respect to CDA). Then, the velocity vector of Cassini in SCNRC is added. This is assumed to be error-less, since for later computations the proper Cassini SPICE kernels are used which do not provide any error estimate and are rather precise compared to the uncertainty of the computed CDA velocities.

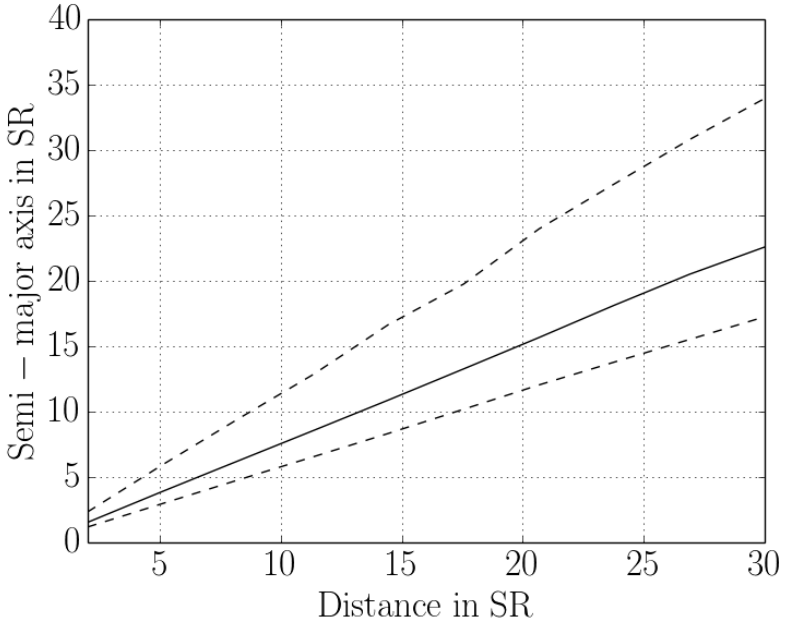


Figure 5.29.: Result of the Monte-Carlo simulation to determine the semi-major axis (Impact Ionisation Detector). The solid and dashed lines show the Median and the IQR of the results vs. the distance from Saturn in Saturn Radius / Radii (SR).

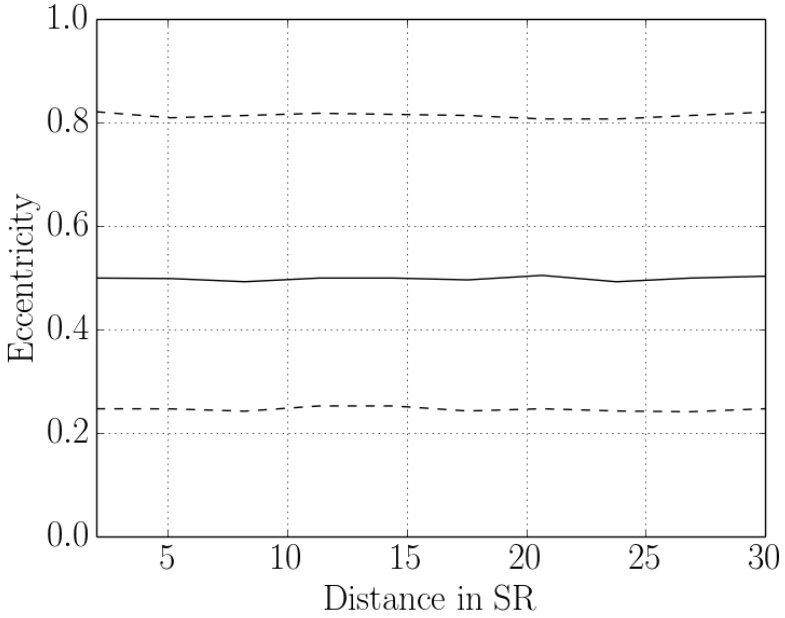


Figure 5.30.: Result of the Monte-Carlo simulation to determine the eccentricity (Impact Ionisation Detector). The solid and dashed lines show the Median and the Inter Quartile Range of the results vs. the distance from Saturn in Saturn Radius / Radii (SR).

With a higher relative Cassini velocity the Median converges to the expected values of a circularly orbiting dust particle. Additionally, the corresponding IQR decreases its width significantly. This shows that the quality of the computed semi-major axis and eccentricity improves for certain Cassini flight geometries. Cassini orbits Saturn in a prograde orbit. Thus, in- and outbound flights to Saturn cause slower impact velocities of circular moving particles with re-

spect to CDA. Since the velocity of Cassini increases with a closer approach to its periapsis, the quality of the determined orbital elements should be better for particles measured closer to Saturn than in the outskirts of the Saturnian system.

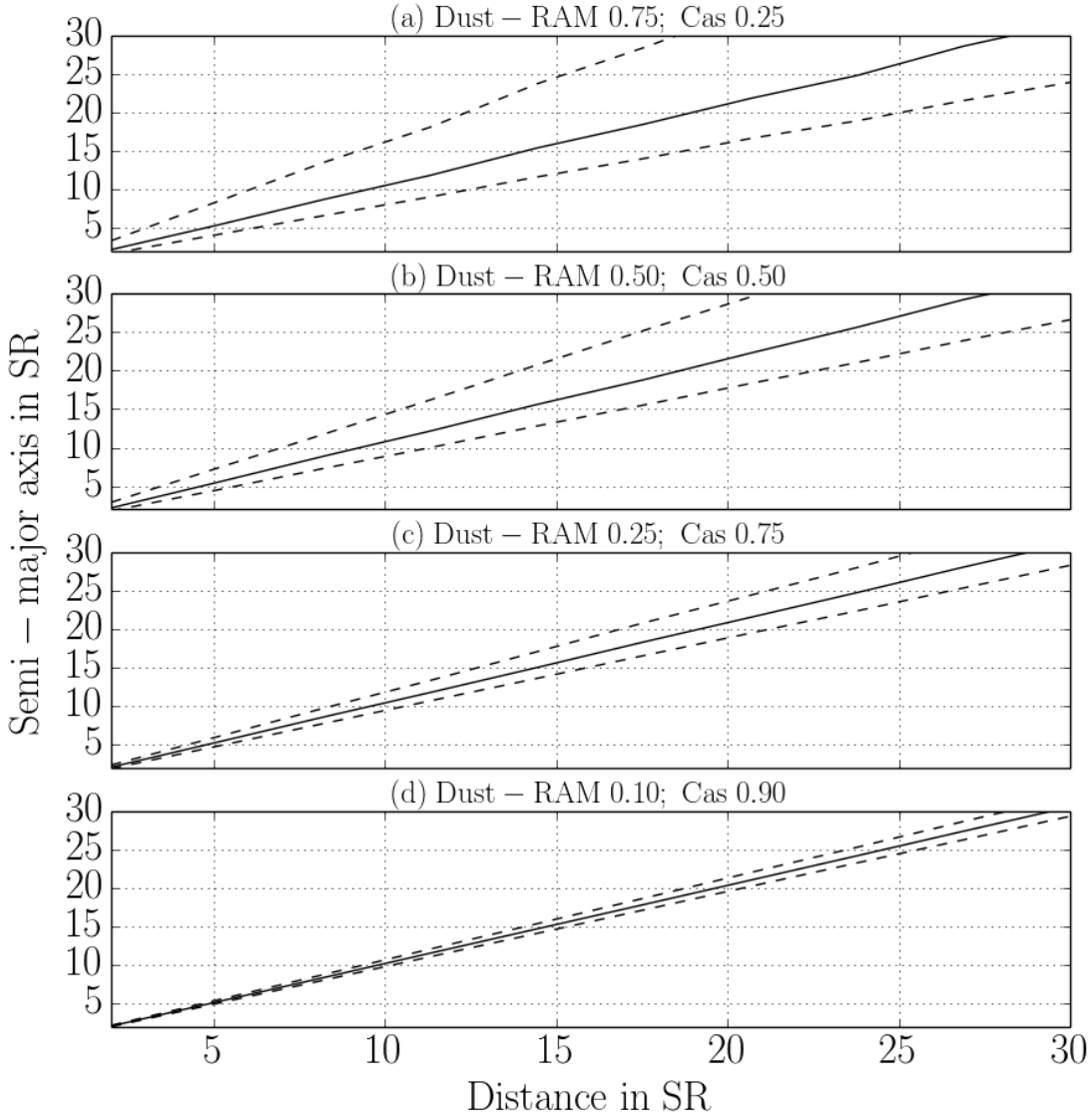


Figure 5.31.: Result of the Monte-Carlo simulation to determine the semi-major axis (Chemical Analyzer Target). The solid and dashed lines show the Median and the Inter Quartile Range of the results vs. the distance from Saturn in Saturn Radius / Radii (SR). Sub-figures a) to d) show different relative velocities of the Cassini spacecraft with respect to the prograde Dust-RAM direction (spacraft moving on a prograde orbit). a) Cassini moves with 25 % Dust-RAM speed, b) Cassini moves with 50 % Dust-RAM speed, c) Cassini moves with 75 % Dust-RAM speed, d) Cassini moves with 90 % Dust-RAM speed.

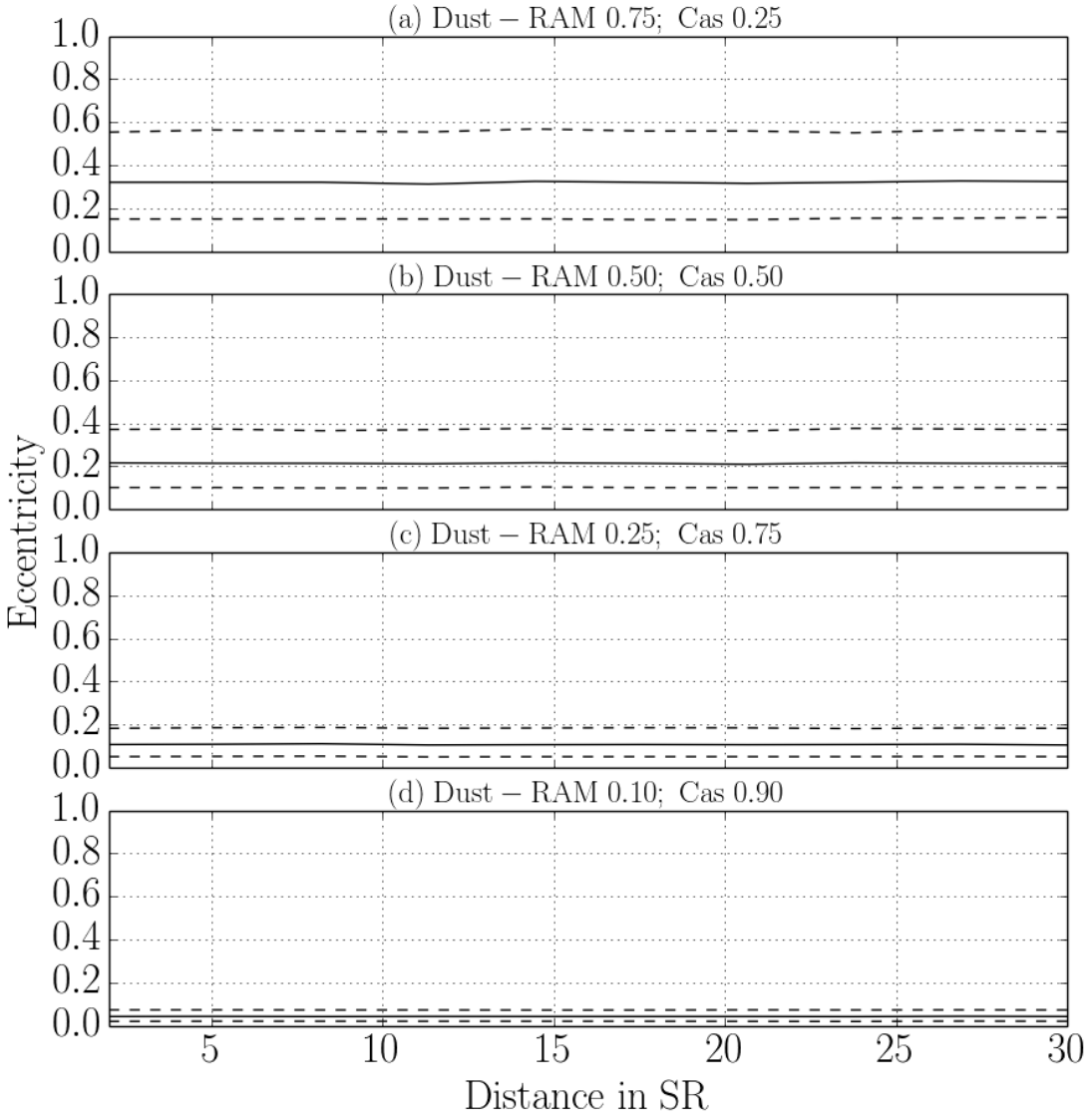


Figure 5.32.: Result of the Monte-Carlo simulation to determine the eccentricity (Chemical Analyzer Target). The solid and dashed lines show the Median and the Inter Quartile Range of the results vs. the distance from Saturn in Saturn Radius / Radii (SR). Sub-figures a) to d) show different relative velocities of the Cassini spacecraft with respect to the prograde Dust-RAM direction (spacraft moving on a prograde orbit). a) Cassini moves with 25 % Dust-RAM speed, b) Cassini moves with 50 % Dust-RAM speed, c) Cassini moves with 75 % Dust-RAM speed, d) Cassini moves with 90 % Dust-RAM speed.

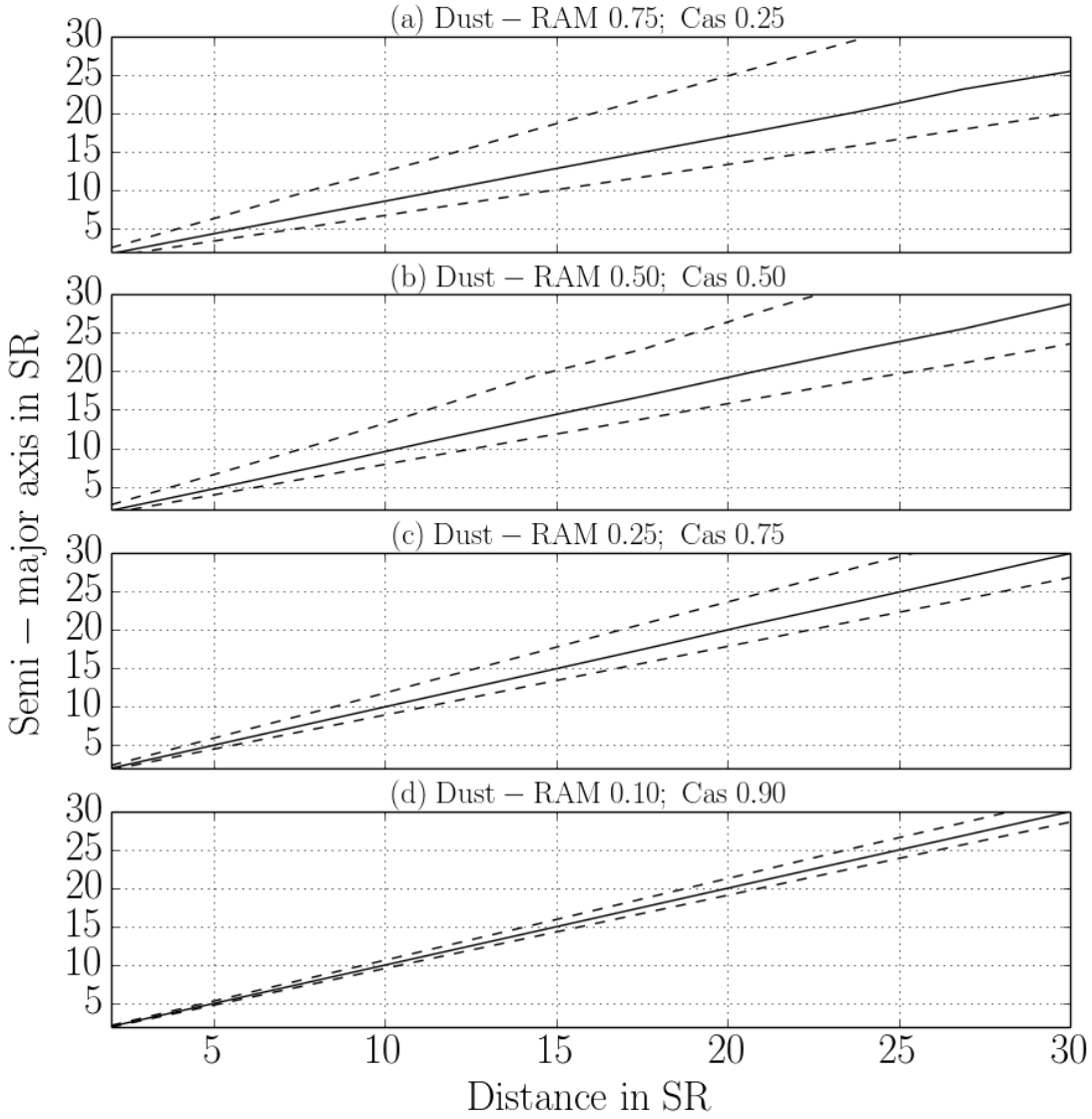


Figure 5.33.: Result of the Monte-Carlo simulation to determine the semi-major axis (Impact Ionisation Detector). The solid and dashed lines show the Median and the Inter Quartile Range of the results vs. the distance from Saturn in Saturn Radius / Radii (SR). Sub-figures a) to d) show different relative velocities of the Cassini spacecraft with respect to the prograde Dust-RAM direction (spacraft moving on a prograde orbit). a) Cassini moves with 25 % Dust-RAM speed, b) Cassini moves with 50 % Dust-RAM speed, c) Cassini moves with 75 % Dust-RAM speed, d) Cassini moves with 90 % Dust-RAM speed.

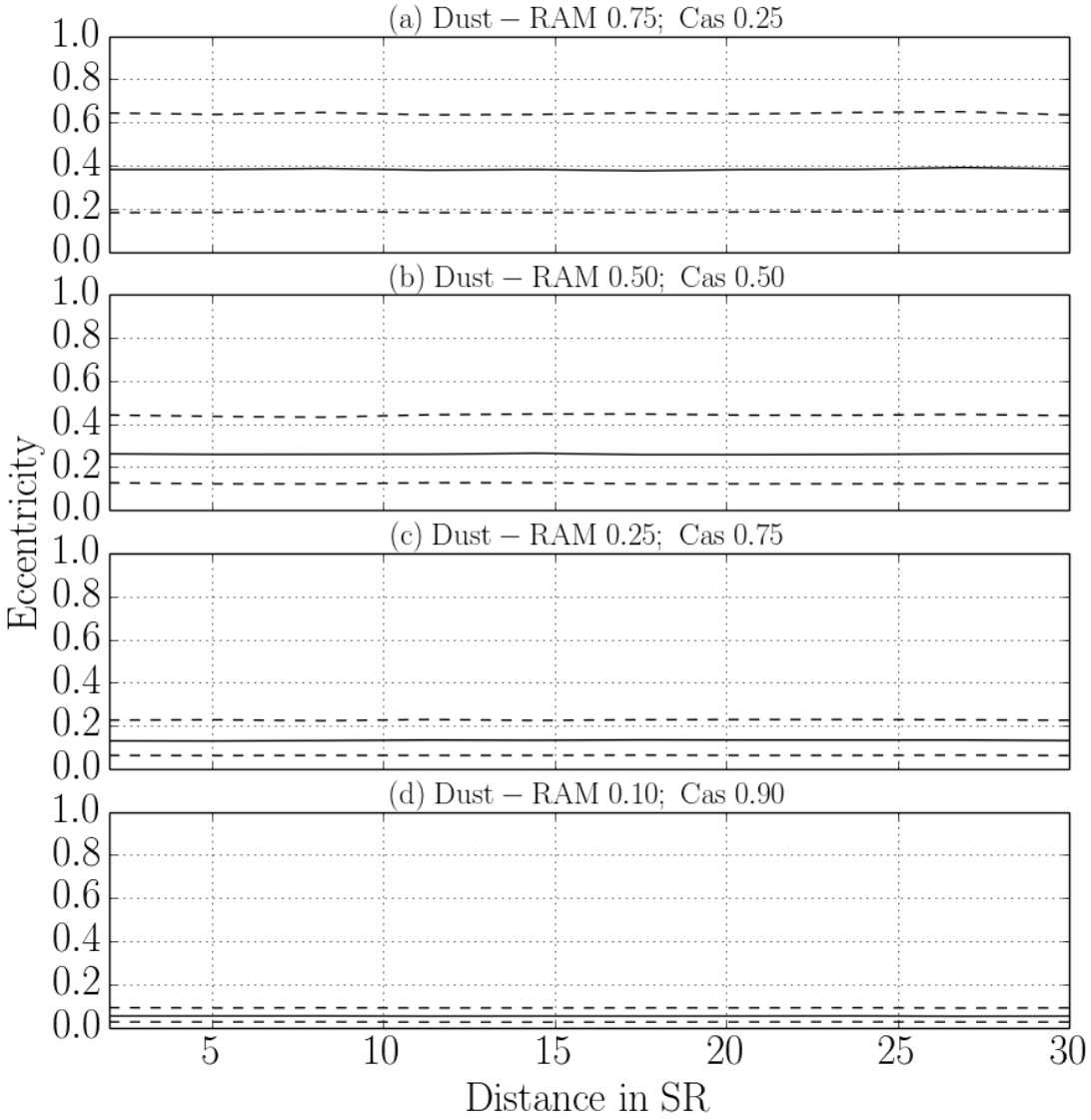


Figure 5.34.: Result of the Monte-Carlo simulation to determine the eccentricity (Impact Ionisation Detector). The solid and dashed lines show the Median and the Inter Quartile Range of the results vs. the distance from Saturn in Saturn Radius / Radii (SR). Sub-figures a) to d) show different relative velocities of the Cassini spacecraft with respect to the prograde Dust-RAM direction (spacraft moving on a prograde orbit). a) Cassini moves with 25 % Dust-RAM speed, b) Cassini moves with 50 % Dust-RAM speed, c) Cassini moves with 75 % Dust-RAM speed, d) Cassini moves with 90 % Dust-RAM speed.

5.2.7. MC algorithm

The final DOCC program has been developed in PYTHON3, although numerical computations perform faster in C or other lower level languages. However, due to the program's complexity and the support of multi central processing units, PYTHON has been chosen. Furthermore the language allows a faster and more agile development. Considering the PEP8 coding standard[4] the developed package can be better read and understand by further developers, for future extensions.

[4] `https://www.python.org/dev/peps/pep-0008/`, 06.04.2018

The developed library contains miscellaneous routines that can be found on the *idust* server. The resulting SQLite database with its tables and entries is vast. A complete list is shown in appendix C.4.

This section explains in each paragraph the computational steps of DOCC.

1. Data preparation The software requires an input file (example see appendix C.3) that contains only data for a particular target, either CAT or IID impacts. The file lists miscellaneous parameters, like detection time, rise time and amplitude, as well as the counter class and several already computed angles between boresight and particular directions (Ram angle, etc.). All data are obtained from the science database Albin [2014]. The following code snippet shows an exemplary SQL query, to obtain data of IID impact events. The listing shows the SQLite shell, after calling the science database [Albin, 2014]. First, the column mode and the header need to be activated. Then, an output file name is defined and finally, the query is submitted to the database. For now, a rather complex query is required that obtains several parameters from different tables that are joined. The linkage is performed with the primary key of the database, namely the Spacecraft Clock (SCLK). Since the counter class schema changed during the mission, the required classes are separated at the SCLK 1496444582. To prevent wrongly determined rise times and amplitudes, both values have to be positive.

```
sqlite> .mode column
sqlite> .header on
sqlite> .output iid_input.txt
sqlite> SELECT ce.SCLK, ce.SCET, ce.cda_counter, printf("%.10f", ce.cda_qi_tr)
    , printf("%.20f", ce.cda_qi_amp), ce.cda_nr_mp_peaks, printf("%.15f", ca.
   cda_bs_x), printf("%.15f", ca.cda_bs_y), printf("%.15f", ca.cda_bs_z),
   printf("%.15f", ca.cda_sun), printf("%.15f", ca.cda_isd), printf("%.15f",
   ca.cda_idp), printf("%.15f", ca.cda_dust_ram_pro_saturn), printf("%.15f",
   ca.cda_dust_ram_retro_saturn), printf("%.15f", ca.cda_saturn), printf("%.15
   f", ca.cda_ring) FROM cda_events AS ce JOIN cda_angle AS ca ON ce.SCLK = ca
   .SCLK WHERE ((ce.cda_counter = 6 OR ce.cda_counter = 7 OR ce.cda_counter =
   10 OR ce.cda_counter = 11 OR ce.cda_counter = 14) AND ce.SCLK >= 1457770000
    AND ce.SCLK < 1496444582 AND ce.cda_qi_tr > 0 AND ce.cda_qi_amp > 0) OR
   (((ce.cda_counter BETWEEN 8 AND 12) OR ce.cda_counter BETWEEN 17 AND 19)
   AND ce.SCLK >= 1496444582 AND ce.cda_qi_tr > 0 AND ce.cda_qi_amp > 0);
```

2. Setting up / calling main routine After the data preparation DOCC requires first a few settings. E.g., the number of seeds as well as the number of central processing units need to be set. Then, the routine can be called and loads the input data as well as functions, SPICE kernels and further constants and parameters. Depending on the target, DOCC computes pre-MC samples. For all following computations, a pre-defined set of directional vectors is computed. The boresight of CDA is aligned in parallel to a chosen Z-axis. Depending on the target's angular PDF, a set of directional, normalised vectors is computed (figure 5.35,

black lines). In the MC simulation, where each particle is considered individually, the boresight vector (blue line) is used and a bisecting line of the angle is computed (green line). The original sample is then rotated by 180° around the green line and points then to the actual pointing direction (gray lines). Then, based on the rise time, a velocity PDF is computed and the resulting velocities are assigned to a vector. So, for every impact event a set of aligned, initial state vectors as seen from CDA is set.

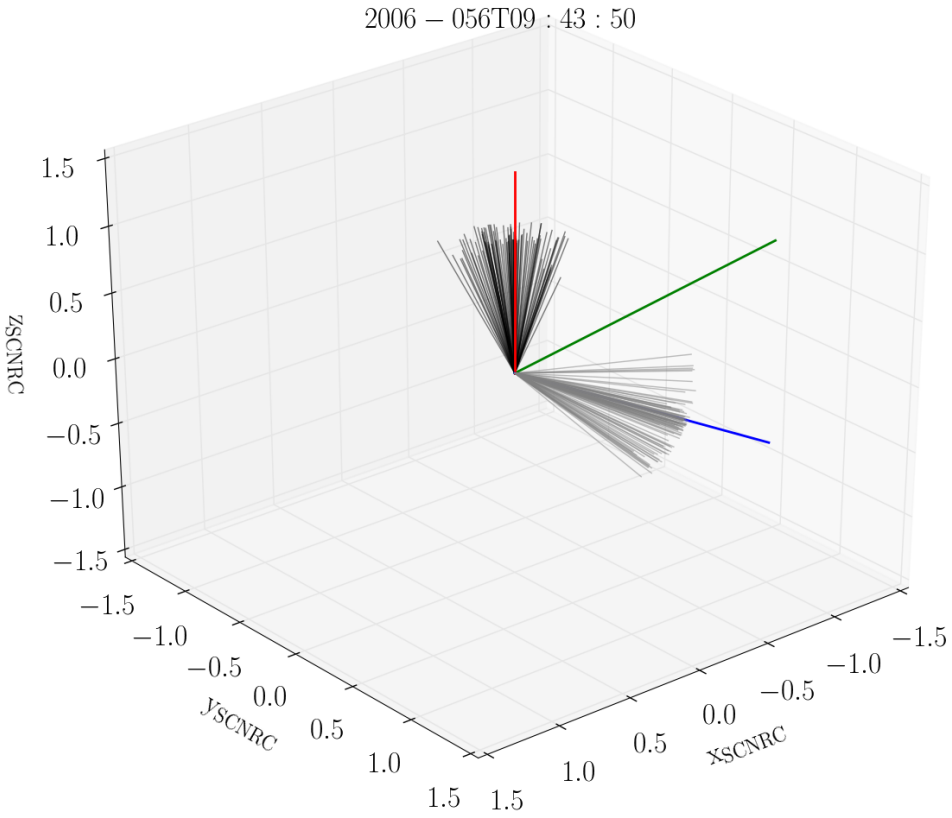


Figure 5.35.: Example how to compute flight direction vectors of an IE. Here, an ensemble of direction vectors is shown based on an IE on the Chemical Analyzer Target at 2006-056T09:43:50. The 3 dimensional plot shows all vectors in Saturn-centric non-rotating coordinate system, where the red line is the Z-axis. The flight direction re-sampling is done in spherical coordinates around the Z-axis (black lines). Based on the boresight of CDA (blue line) the bisecting line of the angle (green line) is computed by adding boresight and Z-axis. Normalized, this vector is used to compute a rotation matrix. The black lines are then transformed to the boresight direction by a rotation of 180° (gray lines). The final ensemble (gray lines) is the solution space of possible dust impacts directions at 2006-056T09:43:50.

In an iterative process, for each impact event each MC state vector is used for further computations. Previous studies have shown that 10,000 re-samples per impact event are sufficient for suitable statistical derivations.

3. Initial state vector determination With the set of initial particle state vectors, first, the detection time corresponding state vector of Cassini is computed. Since DOCC computes only events within the SOI of Saturn, the corresponding state vector is given in SCNRC.

Further, the state vectors of the 8 largest moons are computed. A list of the moons is given in table 5.3. Then, Cassini's distance to the moons is calculated and DOCC checks whether Cassini is within the Hill Sphere of a given moon. The radius of the sphere is shown in the second last column. If the spacecraft is within the gravitational influence sphere of a moon, computation step 4.1 is performed, otherwise the calculation continues in 4.2.

4.1 Moon-centric non-rotating coordinate system (MCNRC) computation During a fly-by at Titan or close encounters with Enceladus (plume science), Cassini was within the Hill Sphere of the moons. Thus, detected particles need to be computed w.r.t. the corresponding body. First, the ensemble of MC re-samples as well as the Cassini state vector are converted in a moon corresponding MCNRC that is similarly defined as the SCNRC. Then, for each re-sample, the orbital elements are computed. An eccentricity smaller than 1 leads to a bound orbit around the moon. Additionally, the periapsis of the particle is checked. If the value is smaller than the radius of the moon (value in brackets in column 4 in table 5.3) the particle was emitted from the moon and is on a ballistic trajectory (falling back to the moon). Note that the bound orbit solutions are unstable due to resonances with Saturn and the other moons. The solutions might be only valid for a short period of time. If the eccentricity is larger or equal 1, the particle is on an escape orbit. In this case three solutions are possible:

In-falling dust In this case the dust is of "moon-exogenic" origin. E.g., interplanetary and planetary dust that crosses the Hill Sphere of the moon.

Emitted dust Due to cryo-vulcanic activities on Enceladus, ice particles are emitted into space. Also impact driven mechanism are possible, like ejecta clouds. These particles are on an escape orbit and are back-propagated in time until they reach the surface of the satellite. Then the corresponding satellite-graphic coordinates of origin are computed.

Fly-by dust The third case describes dust particles that fly-by at the Saturnian moons.

Un-bound fly-by and in-falling dust particles are then computed back until they reach the Hill Sphere of the moon. This is done iteratively and adaptively: The particle is propagated back until it crosses the sphere. Then the integration time is reversed and the computation steps are reduced. This is performed until the integration time reaches 1 second. At the Hill sphere, the time corresponding moon state vector is calculated, and the impact event state vector is transformed from MCNRC to SCNRC.

Table 5.3.: 8 largest Saturnian moons, ordered by the distance from Saturn. The first column is the name of the moon and the second column lists the corresponding Navigation and Ancillary Information Facility ID. a is the semi-major axis (in Saturn radii $r_{\text{sat.}} = 60,268.0$ km). m is the mass in kg. The fifth column shows the radii of the ellipsoidal shape of the moon in x,y and z direction (and mean). r_{HS} is the radius of the Hill-Sphere (from the moon's center). The last column is the ratio $f = m_{\text{Saturn-Moon}}/m_{\text{Earth-Moon}}$ for comparison. From NASA/NSSDC, NASA/JPL [b].

Moon	Naif ID	a [R_s]	m [kg]	$r_x; r_y; r_z; (r)$ [km]	r_{HS} [km]	f
Mimas	601	3.079	$3.751 \cdot 10^{19}$	208; 197; 191; (199)	520	$5.1 \cdot 10^{-4}$
Enceladus	602	3.950	$1.079 \cdot 10^{20}$	257; 251; 248; (252)	949	$1.5 \cdot 10^{-3}$
Tethys	603	4.889	$6.176 \cdot 10^{20}$	538; 528; 526; (531)	2101	$8.4 \cdot 10^{-3}$
Dione	604	6.262	$1.096 \cdot 10^{21}$	563; 561; 560; (561)	3257	$1.5 \cdot 10^{-2}$
Rhea	605	8.745	$2.307 \cdot 10^{21}$	765; 763; 762; (763)	5830	$3.1 \cdot 10^{-2}$
Titan	606	20.274	$1.346 \cdot 10^{23}$	(2575)	52399	1.8
Hyperion	607	24.904	$5.586 \cdot 10^{18}$	180; 133; 103; (139)	2229	$7.6 \cdot 10^{-5}$
Iapetus	608	59.084	$1.806 \cdot 10^{21}$	746; 746; 712; (735)	36290	$2.5 \cdot 10^{-2}$

4.2 SCNRC computation The particle ensemble in SCNRC is handled similarly as described before in MCNRC. The orbital elements of the state vectors are computed and two general solutions are possible: bound and unbound orbits. Bound particles are e.g., E-Ring particles. Besides the orbital elements the Tisserand parameters are computed w.r.t. the larger moons of the Saturnian system. Future scientific analysis that consider this parameter might determine the interaction of dust particles in the E-Ring with Enceladus or with Titan. Unbound particles are:

In-falling dust / Fly-by dust In this case the dust is of exogenic origin.

Nano streams Charged nano sized particles that move away from Saturn. The corresponding impact signals are weak and do not shown strong features, thus, these particle types might be sorted out by the definition range of the used calibration functions.

Particles of exogenic origin are propagated back in time, until they reach Saturn's SOI within an integration time margin. The computational procedure is similar as described before for the moon Hill sphere intersection. The final accuracy is 1 s.

5. Helio-centric ecliptic coordinates (HCEC) computation At Saturn's SOI the entering velocity is calculated as well as the radiants in SCNRC, Saturn-centric rotating coordinate system (SCRC) and ECLIPJ2000. The corresponding SOI intersection time is then required to compute the state vector of Saturn. Afterwards, the MC state vectors of the impact event are converted to ECLIPJ2000, where the orbital elements are determined. An eccentricity smaller 1 indicates an IDP, originating e.g., from a comet or from catastrophic collisions of planetary bodies. Unbound solutions correspond e.g., with particles from the ISD or from other exogenic sources. Further, β-meteoroids are also unbound solutions. Whether this particular particle class can be measured and dynamically determined needs to be studied.

6. Collecting results / Statistics At the end of the MC computation all generated data are gathered: orbital elements, Tisserand parameters, etc. Each dynamical solution is handled separately: Moon-bound, Moon-unbound, Saturn-bound, Saturn-unbound, IDP and ISD solutions. For each data set a set of statistical parameters is derived namely the median, MAD, IQR with its corresponding borders, notches and the whiskers enclosing the confidence interval of solutions (95 %).

Additionally, probabilistic parameters are determined, too. Due to the large FOV and the velocity determination uncertainties an MC run can contain e.g., Saturn-bound as well as Saturn-unbound solutions. Thus, the probabilities are computed of an impact event being Moon-bound, Moon-unbound etc. Also, mass, mean impact velocity and further values are determined.

7. Data storage All derived statistical parameters are stored in an SQLite database that is stored on the *idust* server for further scientific applications. The database contains several tables for each solution space. The SCLK is the unique primary key of the database. A complete list of the database content can be found in the appendix C.4.

5.2.8. MC Test Cases

This section applies the shown computational methods on actual CDA detections. The first section shows an extensive analysis of a single particle and the last section lists some statistics of 1,000 randomly chosen particles from the CDA database [Albin, 2014]. For the analysis only CAT impacts have been selected from the database. 1,000 detections from the database have been extracted, where Cassini was in the E-Ring (distance between 5 to 8 SR) and the CDA boresight direction was close (smaller than 15 degrees) to the prograde Dust-RAM direction. Thus, most particles in the extracted set are detected E-Ring particles.

Test particle Here, a randomly chosen test particle is discussed. The Coordinated Universal Time (UTC) time stamp is 2006-056T09:43:50 corresponding to a SCLK of 1519553623 (primary key of the science database). At this time, Cassini was around 5.6 SR away from Saturn and close to Saturn's ring plane. CAT is the target and the rise time of the QC channel is $t_r = 1.017 \cdot 10^{-5}$ s, which corresponds to an impact speed of $v_{impact} = 3.47$ km/s.

For this test particle 5 MC simulations have been done with 5 different seed sizes, namely 10^2, 10^3, 10^4, 10^5, 10^6. The results for one million seeds are shown in figure 5.36. The four panels show the results of the MC computation where the number of seeds is shown vs. the orbital element range. From top left to bottom right: eccentricity, inclination, semi-major axis, longitude of the ascending node. The solid red and blue lines indicate the mean and median of the distributions. Additionally, the dashed lines show the IQR of the simulated data. The color corresponding numbers above the lines show the mean, median and IQR. The standard deviation and MAD are shown in brackets. The mean and median results deviate although

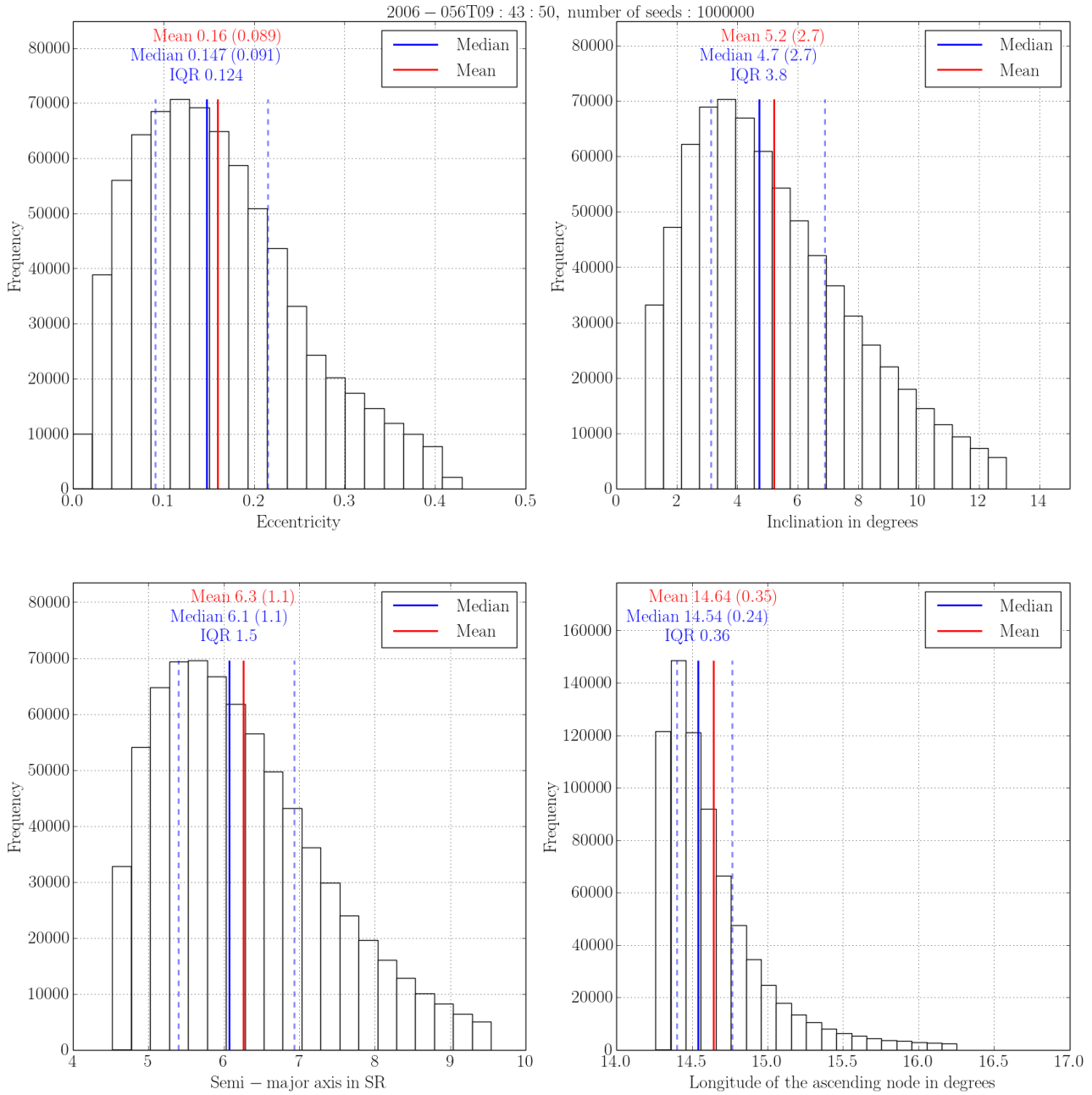


Figure 5.36.: Determined orbital elements of a Chemical Analyzer Target impact at 2006-056T09:43:50. Left top panel to bottom right: eccentricity, inclination, semi-major axis, longitude of the ascending node in Saturn-centric non-rotating coordinate system. Results based on a 1,000,000 seeds Monte-Carlo simulation. The solid red and blue lines indicate the mean and median respectively. The blue dashed lines show the corresponding Inter Quartile Range boundaries. Above the lines the results are listed (top to bottom): mean and standard deviation, median and Median Absolute Deviation, Inter Quartile Range.

a rather large re-sampling size has been chosen. This is due to the non-Gaussian shape of the resulting distributions. Thus, for the final computations, the median and corresponding descriptive statistical parameters are chosen. Also, the median is robust against outliers which might occur randomly in a MC simulation. The results of the simulation with smaller seeds sizes are shown in the appendix C.2. Even a smaller seed size of 10^2, or 10^3 leads to sufficient

results. However one should consider that the impact velocity is rather small, resulting in a narrow velocity PDF. Wider PDFs, which correspond to a high speed impact, need a larger re-sampling size to compute proper orbital elements.

The results (median and MAD in parenthesis) of the shown particle (for a seed size of 10^6) are:

$$e = 0.147 \ (0.091) \tag{5.17}$$

$$a = 6.1 \ (1.1) \ \text{SR} \tag{5.18}$$

$$i = 4.7 \ (2.7) \ \text{deg} \tag{5.19}$$

$$\Omega = 14.54 \ (0.24) \ \text{deg} \tag{5.20}$$

$$\omega = 66 \ (81) \ \text{deg} \tag{5.21}$$

The particle revolves Saturn within the E-Ring on a slight eccentrically shaped orbit with no large inclination with respect to Saturn's ring plane. Also the errors are within a sufficient range for further data interpretation. However, the argument of peripasis shows a large deviation. This is due to the resulting distribution of the MC seeds. In this example most results are close to 0°, thus, the distribution is spread at small angles and large angles close to 360°. Computing the mean or median causes a statistical bias in the computation. Thus, ω (and also Ω) need to be computed the following way:

1. Compute the ensemble of angles
2. Convert the angles in Polar Coordinates with $x = \cos(\phi)$ and $y = \sin(\phi)$, where ϕ is the angle. The resulting vector with the entry x and y has the length 1.
3. Compute the median direction vector and normalize.
4. Compute the angular distance of the median vector to all ensemble vectors with the dot product and compute the MAD.
5. Convert the median vector to an angle with the arcus tangens (since this requires a case-by-case analysis, most programming languages have the function *atan2* implemented).
6. Round the median and MAD angles properly.

This leads to an argument of periapsis of (median and MAD in parenthesis) :

$$\omega = 19 \ (30) \ \text{deg} \tag{5.22}$$

The total central processing unit time for one million particles is approximately 25 seconds.

Ensemble of test particles Here, 1000 test particles have been computed. The database query constrains have already been explained: Cassini within the E-Ring, close Dust-RAM angle, Cassini close to the ring-plane.

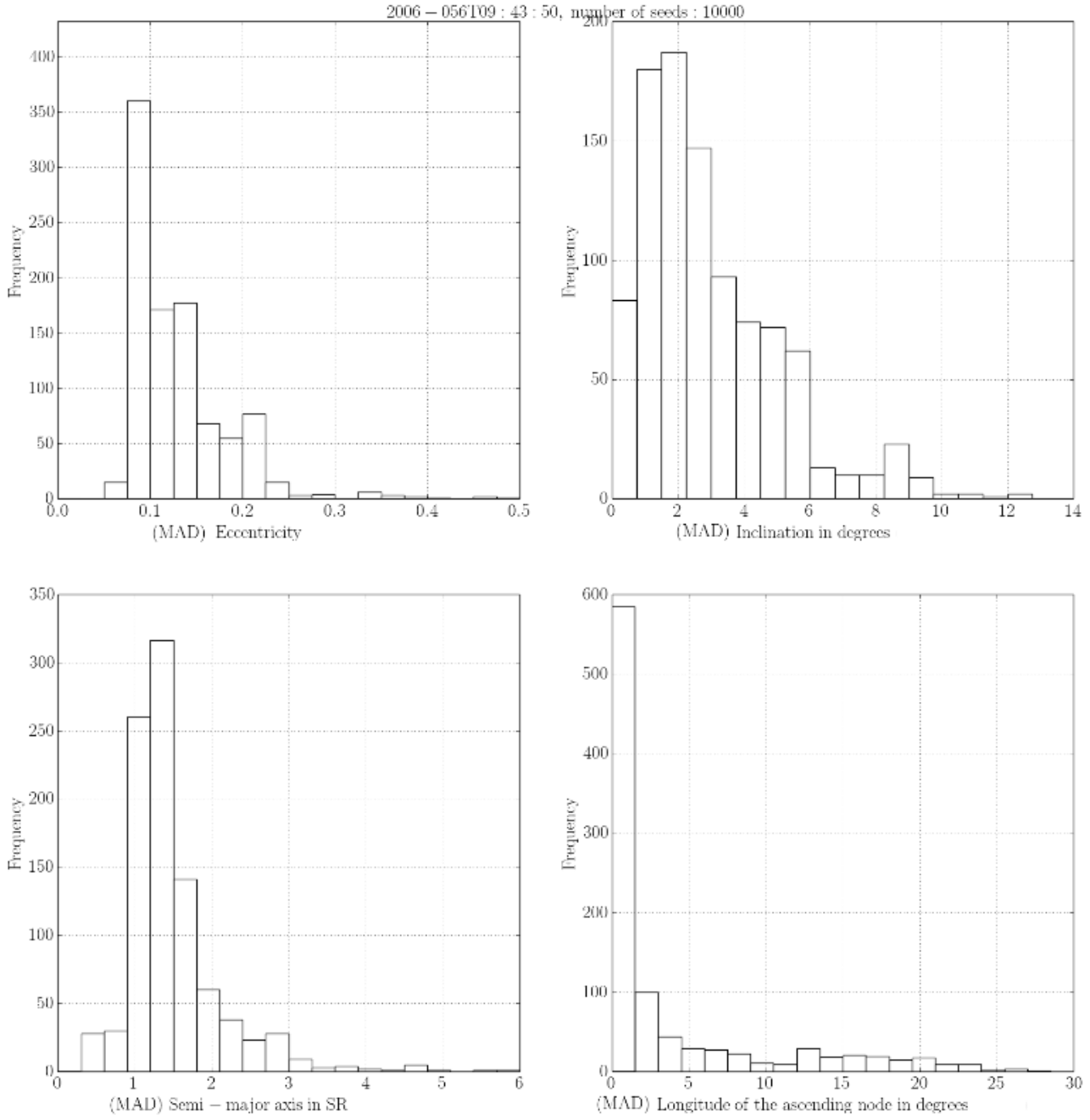


Figure 5.37.: Determined Median Absolute Deviation distribution of the orbital elements of 1,000 randomly chosen test particles on the Chemical Analyzer Target. Left top panel to bottom right: eccentricity, inclination, semi-major axis, longitude of the ascending node in Saturn-centric non-rotating coordinate system. Results based on a 10,000 seeds Monte-Carlo simulation.

Here, the statistics of the randomly chosen particles is shown. Each orbit has been computed by using 10^4 seeds. This section has shown that a seed number of 10,000 is an ideal compromise of computational performance and sufficient data quality. Figure 5.37 shows the MAD distribution of the 1,000 particles (from top left panel to right bottom: eccentricity, in-

clination, semi-major axis, longitude of the ascending node). The distribution of the argument of periapsis, which is not shown, is a rather broad distribution between 0 and 40 degrees and has the largest uncertainty. This is probably a selection bias effect since the chosen particles are on low inclined orbits revolving Saturn on a rather circular orbit. Slight velocity changes have a large effect on this velocity depending parameter.

The plots (figure 5.37) show that the MAD of the inclination and longitude of ascending node is narrowly shaped. The velocity depending parameters eccentricity and the semi-major axis show that the computed particles have sufficient as well as large errors. The error depends on the impacting particle speed, which depends on the particle origin (E-Ring particle, fast IDP or IDP) and Cassini's flight geometry.

5.2.9. Conclusion and Outlook

This section showed the MC based program DOCC to compute astrodynamical properties of CDA impact events. A large SQLite database has been established that allows further details scientific analysis. The next section shows an analysis of the dynamics of the interstellar dust grain candidates as identified by Altobelli et al. [2016].

With the database further research projects will be done like: E- and Phoebe-Ring analysis; Source mapping of dust emitted from Enceladus' tiger stripes or possible dust clouds around larger moons.

Additionally, the identification of interstellar and interplanetary dust particles will be a major scientific project. Bötsch [2018] and Faber [2018] worked on a first dynamical study on these particles. Both used the orbital elements, as well as radiant information to identify possible candidates of exogenic origin. They used the D_{SH} criterion from meteor research to identify possible streams or clusters and linked these candidates with planetary objects beyond the asteroid belt (e.g., Transneptunian Objects). Further, they used so called graphs to visualise the linkage between the particles in radiant plots[5]. They found miscellaneous minor exogenic streams respectively clusters that could be associated with Transneptunian Objects.

5.3. Scientific analysis: Known Interstellar Dust Grains

This paragraph presents astro-dynamical information of the interstellar dust grains that have been found by Altobelli et al. [2016]. This section shows the data processing procedure and analysis to determine the probability and directionality of the 36 ISD candidates listed in the Science paper. The impact velocity determined from the time-of-flight signal (Altobelli & Postberg priv. comm.) is compared with the results from the QC calibration function [Srama, 2009]. With the orbit computation algorithm DOCC the astro-dynamical properties of the particles

[5]Most commonly, graphs are used e.g., in a social media context, where "social graphs" are used to analyse the relations between persons, companies and other entities.

are then computed to verify the interstellar origin by comparing the in-bound radiants with the ISD upstream direction determined by Landgraf [2000]. Furthermore a machine learning algorithm is shown to exclude possible outliers from the data set. Finally, an update is provided of the current interstellar dust upstream direction.

5.3.1. Interstellar dust particles

Altobelli et al. [2016] identified 36 candidates of dust grains originating from interstellar space and the ISD, respectively. For their analysis they used dust grain detections from the small CDA CAT. Thus, for all candidates a mass spectrum has been determined. More than half of these particles have been detected beyond 30 SR. Since the E-Ring dominates the dust environment and measurements in the inner parts of the Saturnian system it is difficult to detect IDP or ISD grains close to Saturn. Also, to optimise and increase the probability of an ISD grain detection the CDA boresight was close to the ISD upstream direction as shown in figure 5.38. The corresponding upstream direction was determined by Landgraf [2000], who used miscellaneous in-situ dust and neutral Helium detectors to compute the ISD.

The relative pointing time is almost constant between 0° and 45°. The relative number of detections however does not correlate with the observation time. The histogram of ISD detections peaks at 0° and drops for higher angular distances.

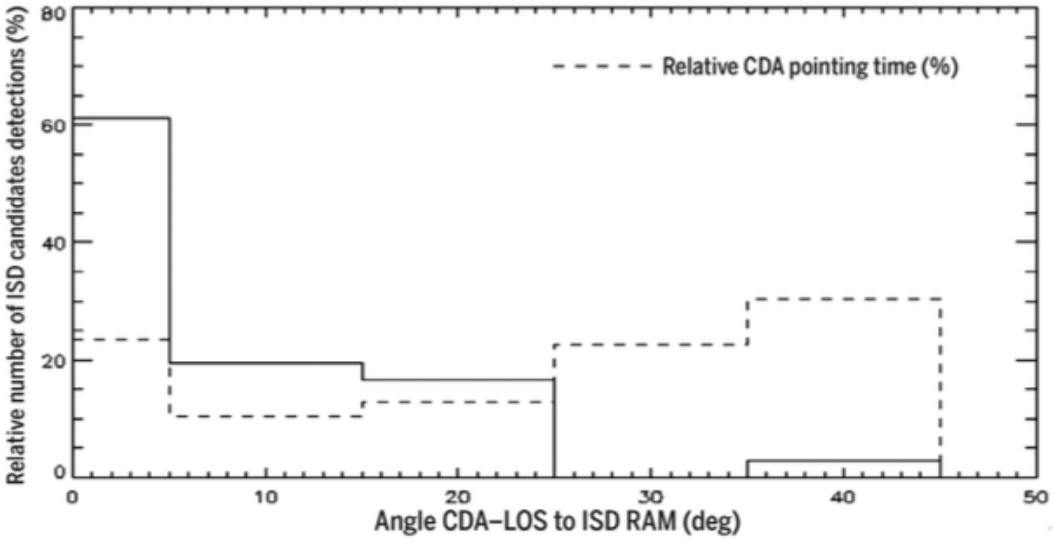


Figure 5.38.: Relative number of interstellar detections (solid) and the Cosmic-Dust-Analyzer pointing time (dashed) to the currently known interstellar upstream direction ($\lambda = 259°$, $\beta = 8°$ in EclipJ2000 [Landgraf, 2000]). Except the first bin, all histogram bins have a width of 10°. From Altobelli et al. [2016].

To verify the interstellar origin of these detections, Altobelli et al. [2016] determined the velocity for each grain with respect to the Saturnian system / center. The velocities have been computed by using the oxygen and carbon cation lines in the spectra as described in the supplementary materials from Altobelli et al. [2016]. Figure 5.39 shows the identification of

the particles, based on the velocity and distance to Saturn in R_s. The solid line indicates the particle speed of grains revolving Saturn on a circular Keplerian orbit. Differently dashed lines above the solid line indicate the injection velocity of theoretical IDPs intersecting Saturn's Hill sphere. Possible interstellar grains are shown as an asterisk above all lines. These measured speeds deny any other dust sources but the interstellar direction.

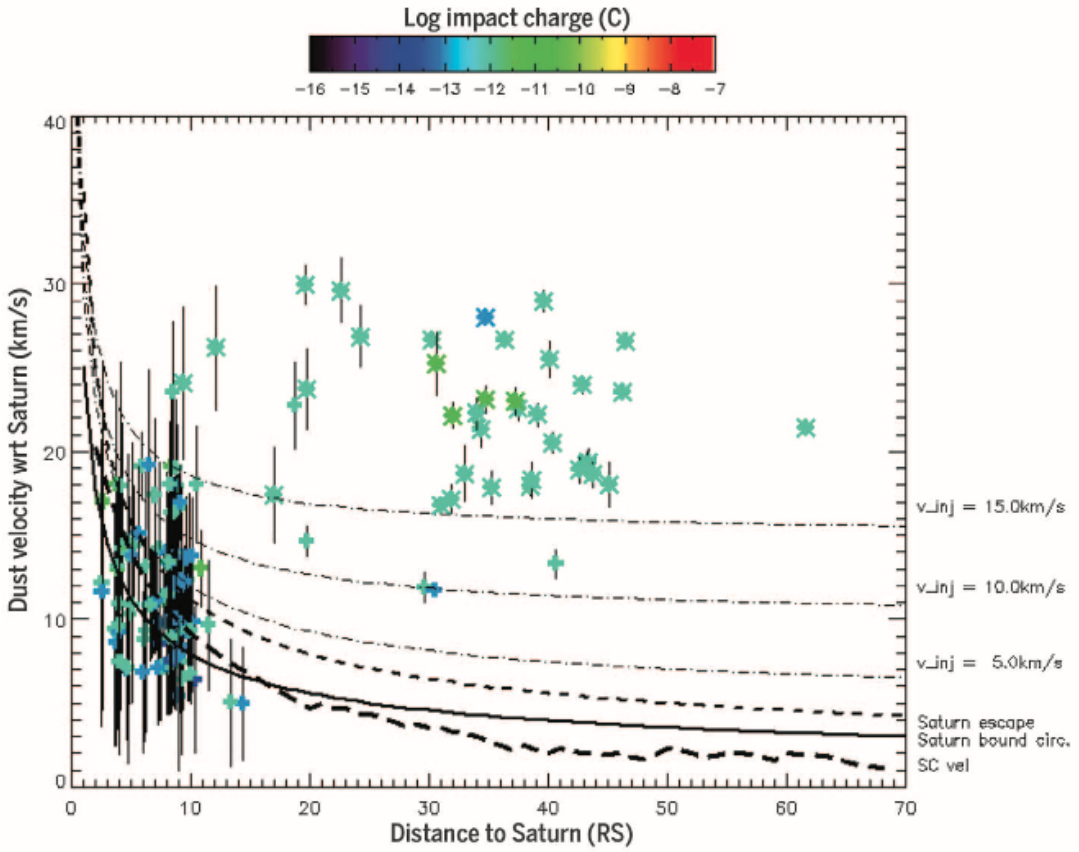


Figure 5.39.: Velocity of interstellar dust grain candidates (asterisk) with respect to the Saturnian center vs. the distance in Saturn Radius / Radii (abbreviated in the plot with RS). The solid line indicates the speed of a dust particle on a circular bound orbit. The dashed lines above indicate theoretical interplanetary particles and their corresponding injection speed at the Hill sphere of Saturn. From Altobelli et al. [2016].

Another possible dust source beyond Saturn are Halley-type comets that revolve around the Sun on very eccentric orbits. Ejected particles from these dust sources could bias the detection likelihood of the ISD candidates, since the eccentricity and corresponding velocity is within the error range of the instrument. However, two observations do not fit with a possible bias from Halley-type comets. First, the spectra did not show any water contribution, as expected from a particle emitted from a comet. Second, dust from these comets is expected to be more directionally isotropic than particles from a predominant interstellar direction. However, as already described in figure 5.38 most candidates have been measured within a close angular distance between the ISD upstream direction and CDA's boresight, although the observation time is constant over a broad angular distance range.

To verify the interstellar origin of these particles, the developed Monte-Carlo code DOCC was used to analyse the directionality of the particles and to re-compute the mean upstream direction of interstellar dust particles.

5.3.2. Analysis

This section describes the data reduction pipeline and the astro-dynamical analysis of the possible ISD candidates found by Altobelli et al. [2016].

Monte-Carlo based probability For the analysis Altobelli & Postberg (priv. comm.) provided a dataset of 35[6] interstellar dust candidates. This dataset contains the impact event corresponding SCLK and the velocity in km/s determined by the time-of-flight signal. These data are shown in the first two columns in table 5.4. Based on the DOCC algorithm the mean impact velocity and corresponding error (standard deviation) is shown in column 3 and 4. The velocities for the DOCC are computed with the signal rise time from the QC channel and the error is determined by a mean velocity depending probability density function as described before in the DOCC section. The resulting DOCC SQLite database is approximately 2 GB large and contains over 300,000 impact events. Miscellaneous tables contain derived astro-dynamical properties and are separated by different possible dynamical solutions like: Saturn-bound orbit, Moon-bound orbit or IDP. A table called *isd* contains the solution of un-bound exogenic particles that have an eccentricity $e \geq 1.0$ in HCEC. In the current version, the table name appears misleading, since particles in this table are not consequently of interstellar origin but could also originate from a Halley-type comet, collisions or due to measurement uncertainties. The last column of table 5.4 shows the probability of a particle being unbound in HCEC. A few rows are greenly marked. Here, the mean velocity from the QC channel corresponds approximately with the value derived from the time-of-flight signal. Additionally, the particles listed in the fourth to sixth row are not present in the database. Possibly these particles did not fulfil the charge and rise time requirements of the Monte-Carlo code as described in the corresponding section.

A distribution of the probabilities in shown in figure 5.40. The histogram shows the results of the computed probabilities based on a re-sampling size of 10,000. Each bin has a width of 0.1 and in total, the plot shows the distribution of 32 particles. The distribution appears bimodal, separating the probabilities in 50 % / 50 % with two peaks at approximately 0 and 1, respectively.

Additionally, figure 5.41 shows the probabilities p vs. the corresponding QC rise time t_{QC} in μs. Due to the signal discretisation a few points overlap in the plot. The rise time is shown logarithmically.

The plot shows a correlation between p and the rise time. The curve progression appears

[6] 1 event was not provided in the requested dataset

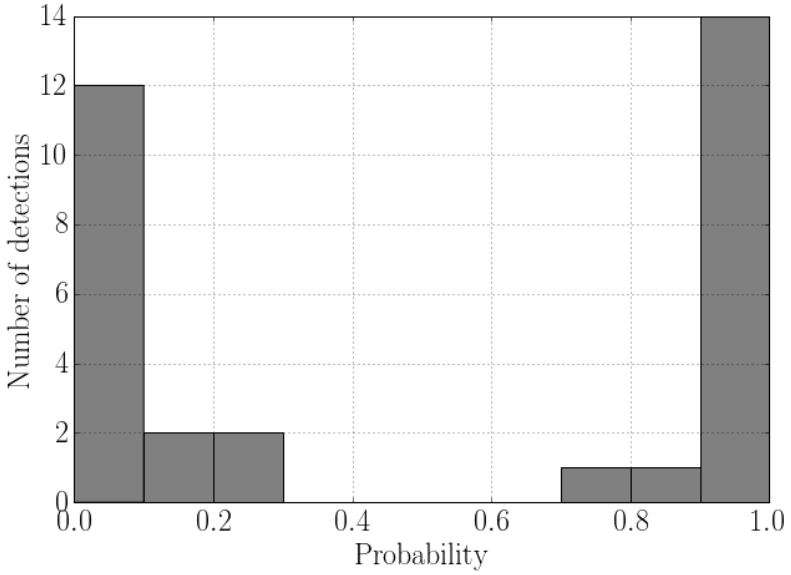


Figure 5.40.: Histogram of the probability distribution of the interstellar dust grains. The probability results from a 10,000 re-sampling Monte-Carlo computation with the Dust Orbit Computation Code (DOCC). Each bin has a width of 0.1. In total, 32 particles are shown.

similar to an invert sigmoid function with a probability close to 1 for very short rise times and a probability close to 0 for larger rise times with a sharp boundary at around 1 μs.

Rise times with $t_{\mathrm{QC}} \lessapprox 1$ μs are almost outside the velocity calibration range. Furthermore, the smallest discretisation step of around 0.4 μs corresponds to a very uncertain speed determination, causing additionally a large velocity error. Due to the resulting high speed, DOCC classifies most particles at the upper calibration velocity range as an HCEC unbound particle. Here, 14 particles with a confidence of at least 95 % are classified as an ISD and have a very small QC rise time.

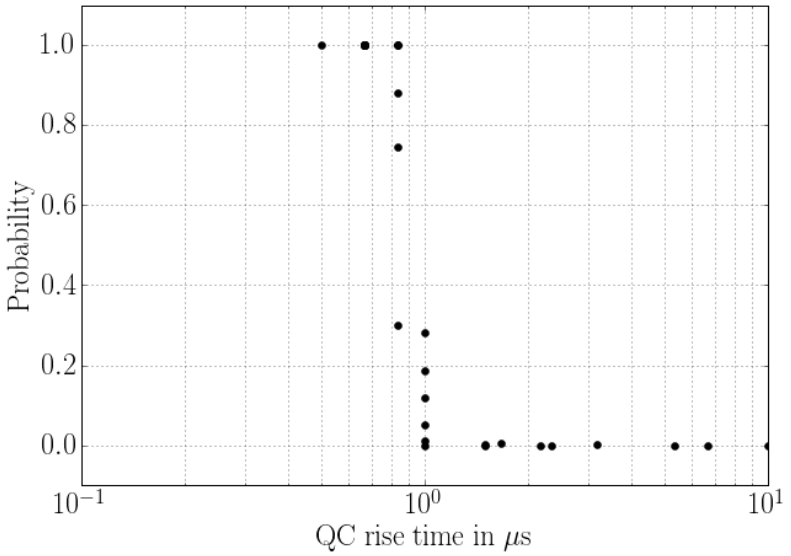


Figure 5.41.: Computed Interstellar Direction probability vs. the corresponding QC rise time. Due to the discretisation of the signal, some points overlap, especially the interstellar dust grains with a probability $p > 0.9$.

Table 5.4.: 35 interstellar dust grains from the supplementary materials of Altobelli et al. [2016]. The first column shows the Spacecraft Clock (SCLK) for each grain and corresponds with the given time stamps in the supplementary materials. The second column lists the lower speed limit of the particle derived from the time-of-flight (TOF) mass-spectrometer (Altobelli & Postberg priv. comm.). Column 3 and 4 show the determined mean and standard deviation velocity based on the rise time of the QC channel. The computed ISD probability, based on the Monte Carlo simulation, is shown in the last column. The green colored lines indicate the particles where the velocity from the QC agrees approximately with the velocity from the time-of-flight signal. These particles are used for the analysis.

SCLK	v_{TOF}	v_{QC}	σv_{QC}	p_{ISD}
1532177344	30	24.6	7.5	0.8805
1577345280	25	30.2	9.2	1.0000
1581719392	25	24.6	7.5	1.0000
1619195939	20	—	—	—
1628199010	30	—	—	—
1634584485	20	—	—	—
1661460235	25	30.2	9.2	1.0000
1664041167	25	40	12	1.0000
1665504393	25	14.6	4.5	0.0000
1667293160	25	20.9	6.4	0.0005
1669487846	20	24.6	7.5	0.2993
1671246161	20	30.2	9.2	1.0000
1672874168	20	14.6	4.5	0.0014
1674800269	25	30.2	9.2	1.0000
1682214182	20	30.2	9.2	1.0000
1684823373	20	24.6	7.5	1.0000
1698470958	20	20.9	6.4	0.2809
1705127968	25	30.2	9.2	1.0000
1709409933	30	20.9	6.4	0.1864
1723435219	30	20.9	6.4	0.0098
1725969537	25	13.3	4.1	0.0055
1726276363	20	20.9	6.4	0.1179
1726334626	20	30.2	9.2	1.0000
1726425331	20	7.8	2.4	0.0007
1726682552	25	30.2	9.2	1.0000
1726740289	20	5.3	1.6	0.0000
1728429362	20	10.7	3.3	0.0000
1728673905	20	20.9	6.4	0.0498
1737028882	25	4.6	1.4	0.0000
1739490759	25	24.6	7.5	0.7441
1748342691	30	30.2	9.2	1.0000

1754251269	25	3.5	1.1	0.0000
1754685693	20	24.6	7.5	0.9997
1754846239	25	30.2	9.2	1.0000
1763786498	20	10.0	3.1	0.0000

ISD directionality data reduction For the astro-dynamical analysis, only the ISD candidates are selected that are within the confidence interval of $p \geq 0.95$. The radiants of the particles are shown in figure 5.42 showing the upstream direction of the dust grains at the SOI of Saturn. The data are shown in ecliptic longitude and latitude, respectively and the spherical coordinate system is distorted to a rectangular map. Black dots represent the median radiant of the MC simulation and the black lines represent the corresponding whiskers of the dataset in latitudinal and longitudinal direction. Whiskers define the border of box plots and have a width of 1.5 times the IQR with respect to the lower and upper quartile, respectively. Assuming a Gaussian distribution the whiskers have a width of approximately 2.7σ.

The radiants cluster approximately at the upstream direction determined by Landgraf [2000]: $\lambda = 259°$ (longitude), $\beta = 8°$ (latitude).

However, 3 radiants appear as possible outliers in the plot. To improve the final median direction computation an outlier detection algorithm is applied on the radiant information.

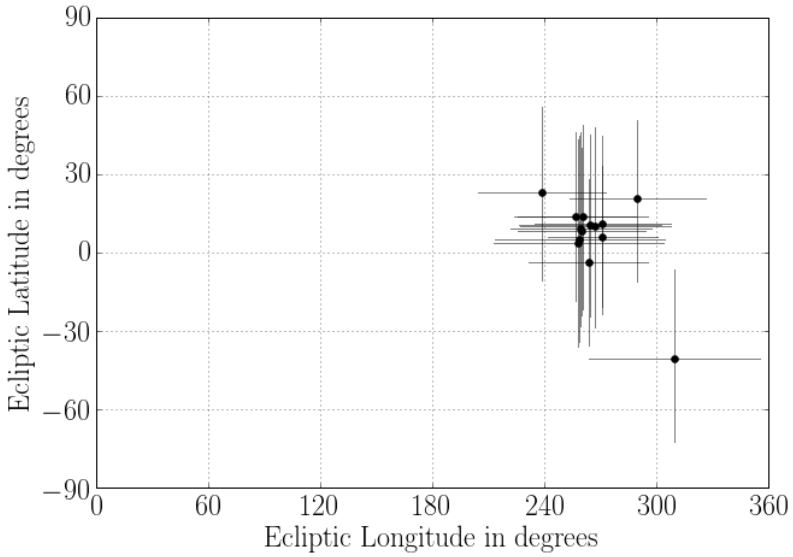


Figure 5.42.: Radiants of the 14 Interstellar Direction (ISD) particles in ecliptic latitude and longitude coordinates. The black dots indicate the median of the Monte-Carlo computation and the corresponding black lines represent the whiskers of the Monte-Carlo solutions.

Covariance matrix and re-sampling Since the number of data points is small and considering the large errors, a proper data preparation needs to be done before applying an outlier detection method. To consider the measurement error, a re-sampling is done, based on the entries in the orbit database. To compute the final statistical quantities, several possible solutions from the MC simulation are stored temporarily. Proper statistics are achieved by using a re-sampling

size of 10,000 per dust grain. To keep the orbit database on a small scale, the re-samples are not saved. However the covariance matrices are computed and stored in the orbit database to allow a simple re-sampling afterwards. These covariance matrices are saved for the orbital elements as well as for the directional information like possible moon source coordinates and radiants. In the database, the radiant covariance matrices are based on three quantities: the ecliptic longitude λ and latitude b, and the intersection velocity v_{SOI} at the SOI. Thus, the quantity vector is:

$$\mathbf{Q} = \begin{bmatrix} \lambda \\ b \\ v_{\mathrm{SOI}} \end{bmatrix} \tag{5.23}$$

Both angles are saved in radians and the velocity is given in km/s. Consequently, the covariance matrix becomes:

$$\mathrm{COV}(\mathbf{Q}) = \left(\mathrm{COV}\left(Q_i, Q_j\right)\right)_{i,j=1,2,3} \tag{5.24}$$

This matrix is saved in table *isd* as a blob named *radiant_ isd_ eclip_ cov*. The following code snippet shows briefly how to obtain the covariance matrix and how to convert to a proper array format in PYTHON for further analysis.

The minimum example needs the packages SQLITE3 and NUMPY, a connection to the orbit database and a set cursor:

```
import numpy as np
import sqlite3

con = sqlite3.connect('cda_orbit.db')
cur = con.cursor()
```

Now, an SQL command needs to be send to obtain the covariance matrix. Note, the complete *WHERE* statement has been simplified.

```
cur.execute('SELECT radiant_isd_eclip_cov FROM isd WHERE ...')
temp = cur.fetchall()
```

The array *temp* contains now the covariance matrices in a string / blob format. In this example, the first entry of *temp* looks like:

```
1 '[[ 0.05897348  0.00112192  0.14228301]\n [ 0.00112192  0.05176816
     0.01958121]\n [ 0.14228301  0.01958121  0.44136885]]'
```

Now, the blob needs to be parsed into a proper format. A code snippet is shown in the following list that can be applied also on other matrices. First, the final array needs to be defined *cov*. In a for-loop, each covariance blob is parsed and appended to this array. In a first step the characters ' \n', '[' and ']' are replaced by an empty character. The covariance is a

3×3 matrix during the MC computation. After converting the array to an SQL blob the line breaks (separation of the individual matrix' rows) and are converted to a string character as well as the brackets.

Afterwards the resulting string contains all data that are only separated by an empty space. Splitting this string results in a 1×9 array. To generate the proper matrix format, the array is reshaped with NUMPY's *reshape* function. In a final step, only the first 2 rows and columns respectively are saved into the final covariance matrix. This sub-matrix contains only the covariances of the ecliptic longitude and latitude, respectively.

```
1 cov = []
2 for k in temp:
3
4     cov_temp = k.replace('\n', '').replace('[', '').replace(']', '')
5     cov_temp = np.array(temp.split(), dtype = 'float').reshape(3,3)
6
7     cov.append(cov_temp[:2, :2])
```

With the obtained longitude and latitude median values from the database a re-sampling can be done. Note: Statistical quantities (median, MAD, etc.) are stored in degrees. The entries in the covariance matrix are computed based on the corresponding radians information. To properly re-sample the values the medians need to be converted to radians. Then, the re-sampling can be done and the final solutions can be re-converted to degrees for further analysis or plotting.

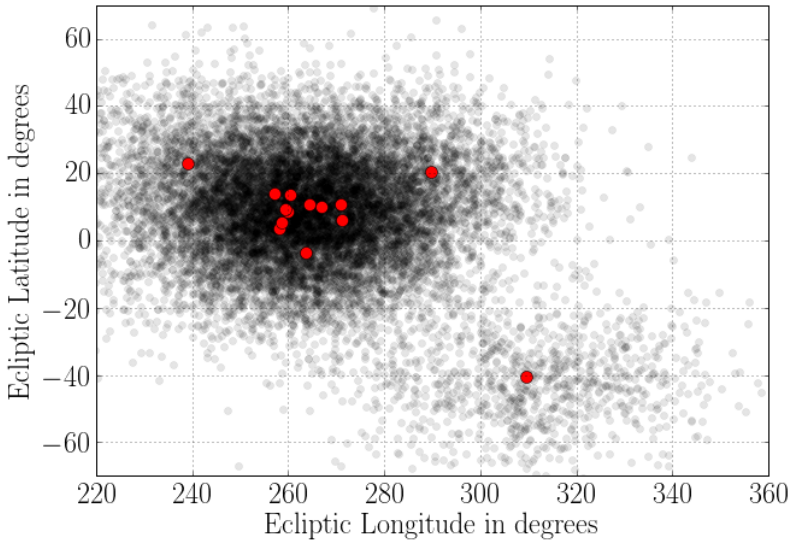


Figure 5.43.: Re-sampled radiants based on the covariance matrices for the 14 interstellar candidates. The median radiants are shown as red dots in ecliptic coordinates and the black dots are the re-sampled ones. For each radiant point a Gaussian re-sampling size of 1,000 has been chosen.

The a-posteriori distribution of the ecliptic longitude and latitude is unknown. Due to CDA's symmetric opening a Gaussian distribution might approximate the MC results. However, a Gaussian distribution is defined on complete $\mathbb{R}$ and CDA's smaller target has an angular observation cone of approximately 60°. Possible outliers may occur that can be neglected since

a re-sampling size of 1,000 per radiant is chosen for the analysis. Note: A Gaussian re-sampling can only be applied close to the ecliptic equator. High latitude values need to consider the Haversine metric[7].

The resulting radiant solution can be seen in figure 5.43. The median and re-sampling radiants are shown as red and black dots, respectively.

Outlier Detection Algorithm Based on the re-sampled radiant data an outlier algorithm needs to be chosen. As already described previously, for this purpose the package SCIKIT-LEARN is used [Pedregosa et al., 2011]. To apply a proper algorithm some properties need to be determined. Based on the problem and corresponding properties, a suitable algorithm can then be chosen and applied. The following list shows some characteristics:

Dimensionality: The problem considers only the ecliptic longitude and latitude. Thus, the problem has a dimension of 2.

Clusters: Landgraf [2000] determined one ISD upstream direction. Thus, only one cluster (no bimodal distribution or higher orders) is present.

Symmetry: The cluster has approximately a rotation symmetry.

Separability: We assume a cluster around the upstream direction with an increasing outlier probability proportional to an increasing angular distance to that direction.

Based on these constrains one algorithm appears suitable for this problem: The *Elliptic Envelope* [Rousseeuw and Driessen, 1999, Hubert and Debruyne, 2009]. Here, the algorithm from the SCIKIT package is used [Pedregosa et al., 2011].

The algorithm is based on fitting a 97.5 % tolerance ellipse on a Gaussian distributed dataset. This ellipse is based on all column vectors $\mathbf{r}$ whose distance equals to the square-root of the 0.975 quantile of the Chi-squared function in d dimensions (for the shown problem: $d = 2$):

$$\sqrt{\chi^2_{d,0.975}} \tag{5.25}$$

The corresponding distance metric is the Mahalanobis distance and is defined as, where $\boldsymbol{\mu}$ is the corresponding mean and C is the covariance matrix:

$$d_{\text{Mahalanobis}(\mathbf{r},\mu)} = \sqrt{(\mathbf{r}-\boldsymbol{\mu})^{\mathrm{T}} \cdot C^{-1} \cdot (\mathbf{r}-\boldsymbol{\mu})} \tag{5.26}$$

[7]The Haversine metric is the metric in spherical coordinates. The distance between two points on a sphere are determined by computing the great circle distance

Highly scattered data or data that are (unknowingly) affected by e.g., a random noise are fitted with a too broad, biased or shifted tolerance ellipse. This leads to a wrong interpretation of the truly underlying distribution and considers outliers as inliers. To compensate the effect of present outliers Rousseeuw and Driessen [1999] introduced an algorithm that replaces the covariance matrix with a robust substitute called *Minimum Covariance Estimate* covariance matrix. The mean values are also replaced by the corresponding *Minimum Covariance Estimate* mean values. A brief description and derivation of these parameters is given in Hubert and Debruyne [2009]. This method considers also a user-given "contamination" factor and fraction, respectively. This value ranges between 0.0 and 1.0. This method can be applied on data where the number of samples exceeds 5 times the dimensionality of the data. In this case, the sample size is 14,000 and the dimensionality is 2, thus the requirement $14,000 > 10$ is fulfilled.

Figure 5.44 shows the re-sampled data and an elliptic envelope (orange area) assuming a contamination fraction of 0.2. 13 median radiants are within the envelope. However, since the contamination rate is unknown different contamination factors are used to determine the fraction of re-sampled inliers per median radiant.

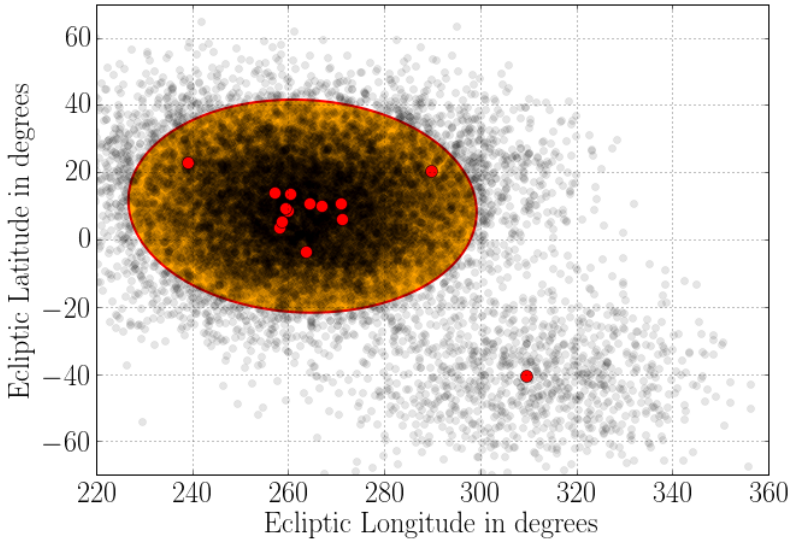


Figure 5.44.: Similar figure as already explained in figure 5.43. The red elliptic line and orange area respectively, include all inliers in the data set assuming a contamination of 0.2.

The inlier fraction vs. contamination rate dependency is shown in figure 5.45. The red and blue curve correspond to the inliers of the re-sampled radiants with a median ecliptic latitude of $\beta < -20°$ and $\beta > 20°$, respectively. The remaining inliers are shown as black curves. Even for a small contamination rate of 0.1 only the black curves have a inlier fraction larger than 95 % while the blue and red curves are significantly outside the cluster of the 11 close radiants. For higher contamination rates, 3 individual clusters are present. Thus, for the final ISD upstream direction determination the inner 11 radiants and corresponding re-samples are used.

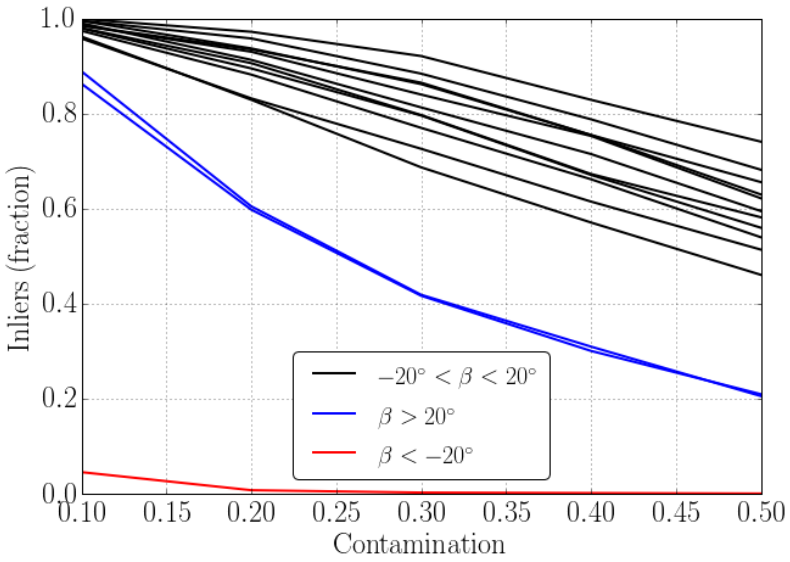


Figure 5.45.: Fraction of inliers within the elliptic envelope vs. the contamination rate. Data from the close cluster are shown in black, the results for the two radiants with an ecliptic latitude β larger than $20°$ are shown in blue and the value with $\beta < -20°$ is shown in red.

ISD upstream direction To determine the upstream direction of the interstellar dust the re-samples of the inner 11 radiants are used to determine a PDF. Each re-sample value is replaced by a Gaussian kernel. All kernels are then summed up and normalized. The resulting distribution is a color map in figure 5.46. The highest probability is indicated as a green triangle at $\lambda = 263°$ and $\beta = 9°$. As a reference, the upstream direction determined by Landgraf [2000] is shown as a blue triangle. Furthermore, a Density-Based Spatial Clustering of Applications with Noise algorithm is applied on the re-sampled data, too. Also this method verified the newly computed upstream dust direction, based on the interstellar dust candidates. However, considering the rather large probability density function, one cannot assume a temporal change of the interstellar upstream direction. Other effects, like the Solar cycle variation are also not taken into account. Missions with the dedicated objective to detect interstellar grains could used these methods to properly clean the measurements and determine proper probability functions over time.

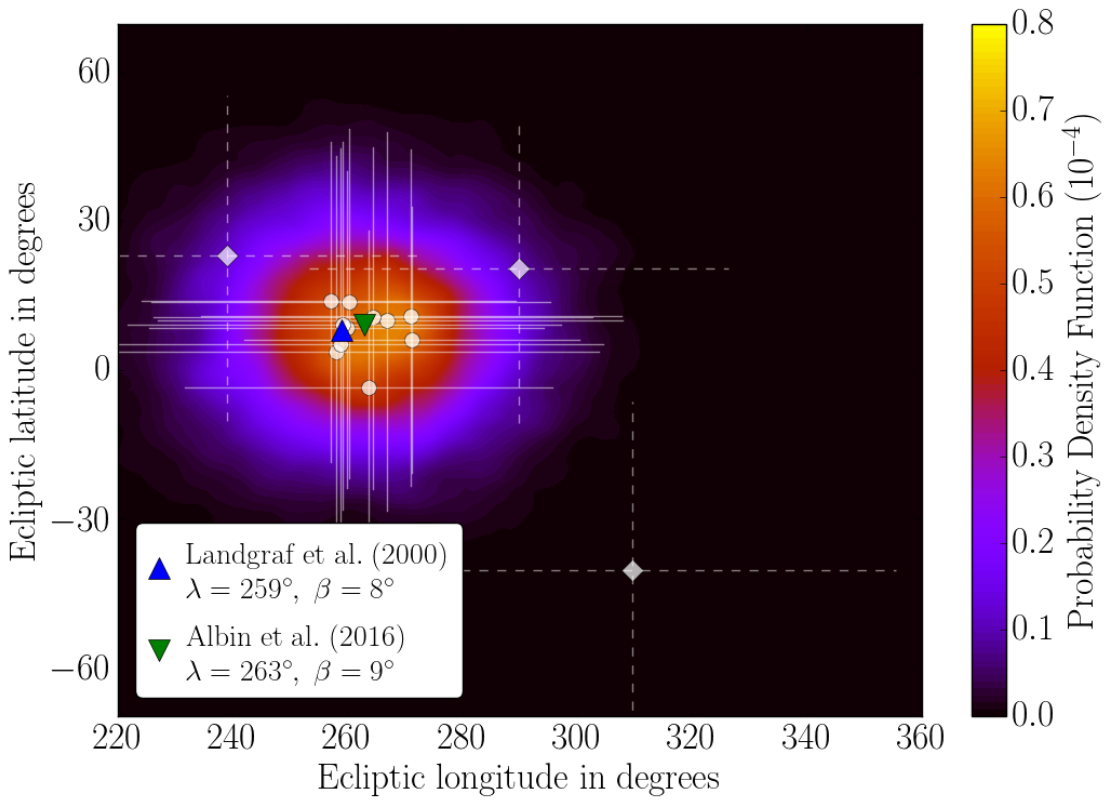


Figure 5.46.: The interstellar dust upstream direction determination based on the 11 radiants (circles) and the corresponding re-samples. 3 outliers have not been considered (diamonds). The color map shows the probability density function of the upstream direction and a green triangle indicates the highest probability. A blue triangle shows the upstream direction as determined by Landgraf [2000].

5.3.3. Astrodynamical Analysis

This section analyses the astro-dynamical properties of the 11 ISD candidates in HCEC. First, the orbital elements and the corresponding errors for three elements are shown. A trajectory visualisation is shown afterwards.

Orbital Elements Figure 5.47 shows the box plot results of the perihelia for each dust grain. The X-axis corresponds to the detection time when the particle has been detected. Each box includes 50 % of all solutions and the inner horizontal line indicates the median. The whiskers (dashed lines) have a (maximum) length of 1.5 times the width of the box.

8 particles have a median perihelion distance between 8 and 10 AU with a 50 % confidence within an interval of approximately 0.3 and 1 AU. 3 particles have a perihelion distance that is closer to the inner parts of the Solar System. The last particle has a median at 6 AU. However, possibly due to the velocity determination error the corresponding uncertainty is high.

Figure 5.48 shows the corresponding eccentricities. Since the eccentricity is velocity dependent the error of the eccentricities adapts accordingly. 10 particles have a median eccentricity between 2 and 5. However, one outlier appears at approximately 12. This dust grain has a QC rise time of 0.5 μs respectively an impact speed of around 40 km/s.

The last figure 5.49 shows that the ISD particles were possibly on a retrograde orbit. Due to CDA's large angular FOV the error-bars appear rather larger.

Due to the high velocity determination error and the large FOV of CDA, the argument of periapsis has a rather larger error up to tens of degrees. However, for the visualisation of the orbits, the median values are used.

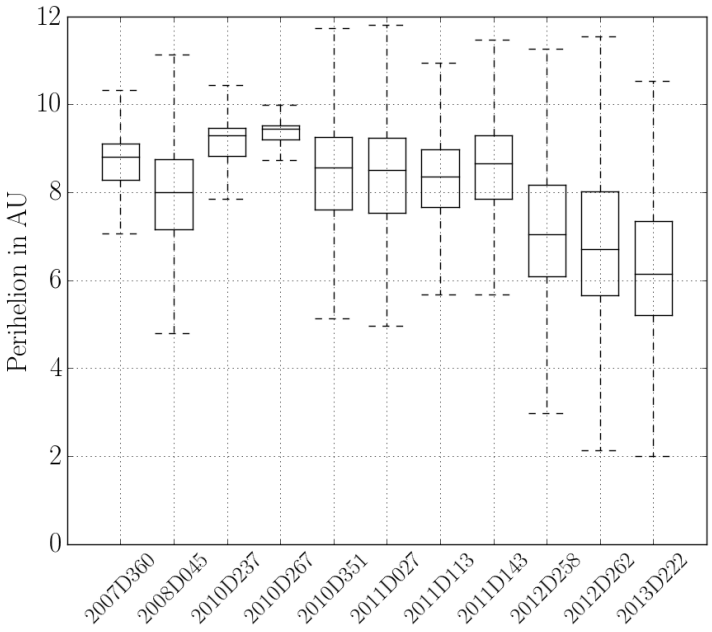


Figure 5.47.: Box plot data visualisation of the perihelia of the 11 Interstellar Direction particles. The perihelia are given in Astronomical Unit (AU) and the corresponding detection time is shown on the bottom axis. The date shows the year and the day of year.

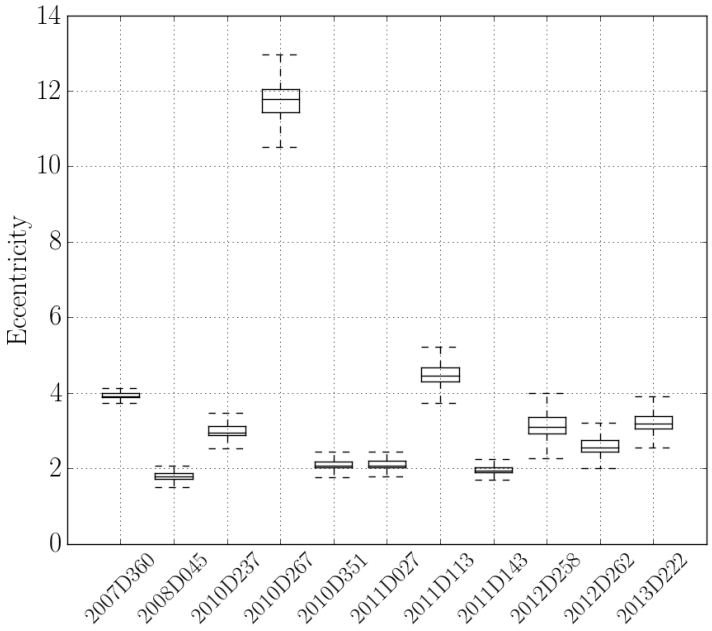


Figure 5.48.: Box plot data visualisation of the eccentricity of the 11 Interstellar Direction particles. The corresponding detection time is shown on the bottom axis. The date shows the year and the day of year.

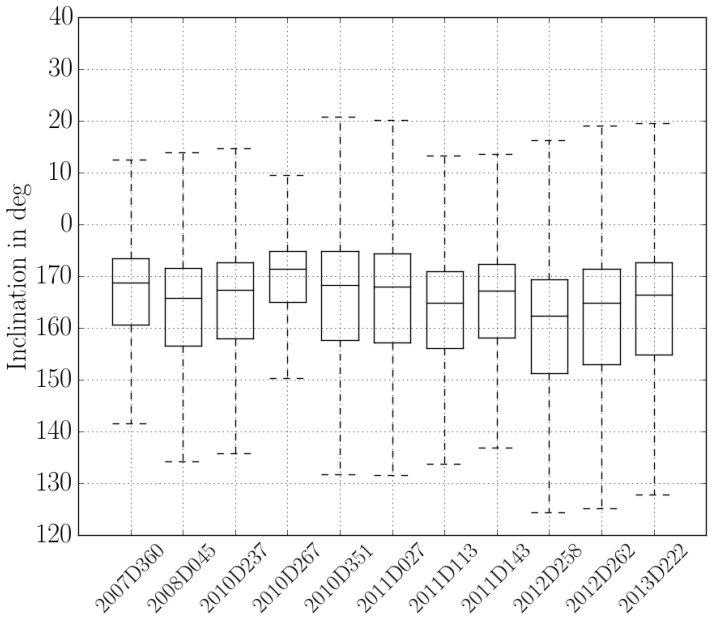


Figure 5.49.: Box plot data visualisation of the inclintation of the 11 Interstellar Direction particles. The inclination are given in degrees and the corresponding detection time is shown on the bottom axis. The date shows the year and the day of year.

Median Trajectories For the visualisation of the orbits the median values from the MC simulation are used. Thus, the directionality of the particles is mainly given by CDA's boresight. To visualise all particles the Kepler elements corresponding covariance matrix has to be used as explained in the section before.

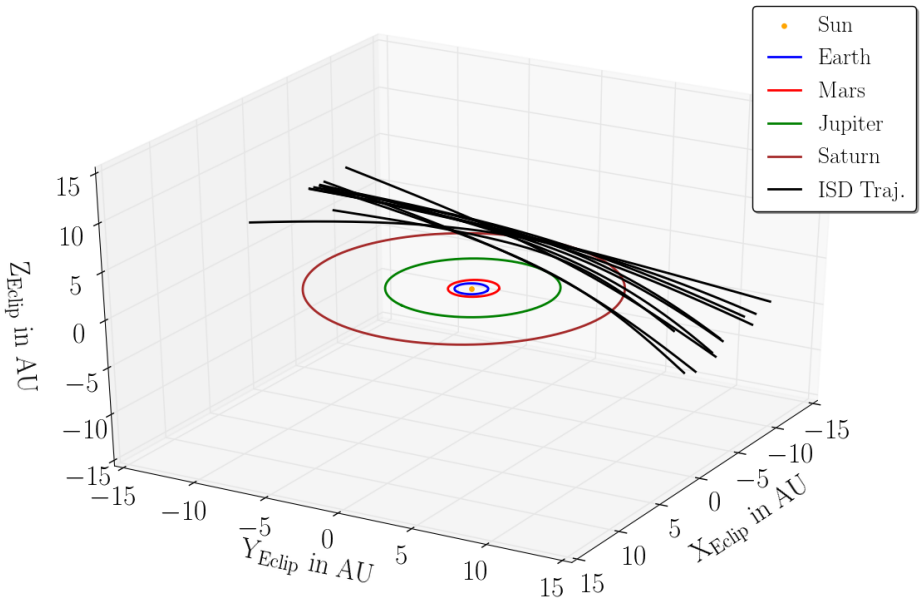


Figure 5.50.: 3-dimensional view of the 11 interstellar median trajectories (black lines) in Helio-centric ecliptic coordinates (HCEC). The upstream direction of the particles is at the top of the figure. Additionally, the orbits of the Earth, Mars, Jupiter and Saturn are shown, too, as well as the Sun in the center.

Figure 5.50 shows the median trajectories of the interstellar grains (black lines) in HCEC.

The Sun is displayed in the center with Earth's, Mars', Jupiter's and Saturn's orbit. The upstream direction of the particles is in the top left. Considering also the topview plot 5.51 (upstream direction: bottom), it can be seen that 8 particles have a well matching trajectory while 3 particles have slightly different upstream directions. Since the orbital element errors are rather large, one can assume that all particles are astro-dynamically linked.

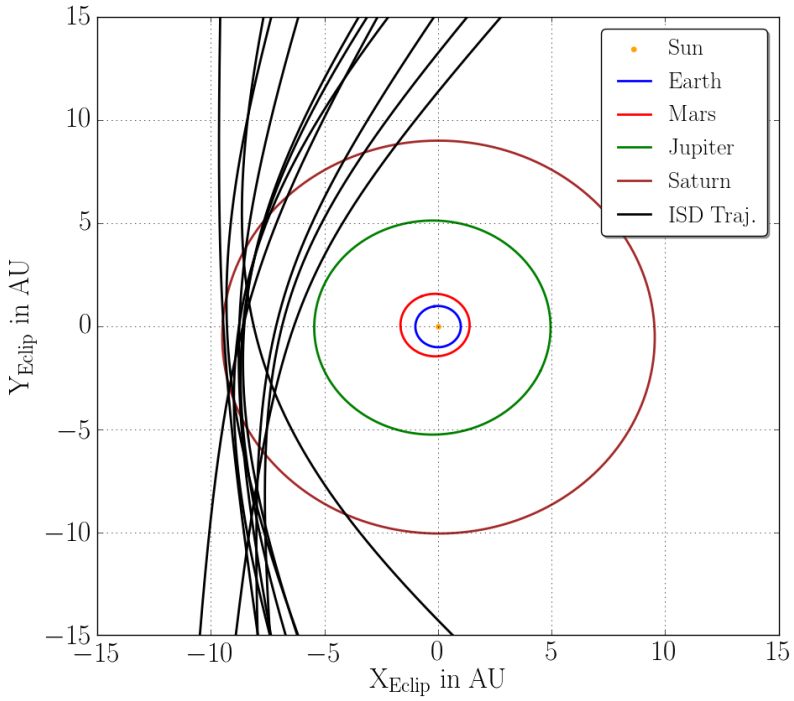


Figure 5.51.: Topview of the 11 interstellar median trajectories (black lines) in Helio-centric ecliptic coordinates (HCEC). The upstream direction of the particles is at the bottom of the figure. Additionally, the orbits of the Earth, Mars, Jupiter and Saturn are shown, too, as well as the Sun in the center.

5.3.4. Summary & Conclusion

This section analysed the astro-dynamical properties of the ISD particles that have been found by Altobelli et al. [2016]. They found 36 candidates with an according spectrum and focused on the chemical composition of the interstellar dust. With the orbit database that has been generated with the Monte-Carlo based DOCC the dynamics of 35 particles have been analysed. Altobelli et al. [2016] used the mass spectra to determine a precise impact speed while DOCC used the velocity determination formulae based on the QI (IID) and QC (CAT) rise time [Srama, 2009]. In a first approach, only particles were considered whose velocities determined by Altobelli et al. [2016] and DOCC match approximately. Furthermore, only grains were used for further analysis that have at least a 95 % confidence of being a dust grain from interstellar space. This lead to 14 particles. An outlier detection algorithm determined a significantly different directionality of 3 ISD grains based on the radiant information at Saturn's SOI. In the last section, some astro-dynamical properties were shown of the remaining 11 particles that appear dynamically linked.

In a future work, the directionality information of the particles can be used to determine further ISD candidates, unbound particles that might emerge from the Oort Cloud or (interstellar) streams.

5.4. NN based velocity calibration

The computed orbit database is based on calibration functions performed by Srama [2009]. He determined miscellaneous calibration functions for the velocity and mass of an impact event. For the mass data he applied hyperplane fitting routines while the velocity functions are polynomial functions, with rather large uncertainties.

This section describes a new regression approach based on NNs. First, the calibration data set is described in detail, to understand its structure that allows to derive requirements for an NN approach. Then, the model is described, and how the model is trained. Afterwards a verification section checks the properties of the models and compares the results with previous calibration functions. Furthermore, a definition range is determined, where the new regression models perform as well as the functions by Srama [2009]. In the corresponding appendix are several results listed from the different channel models.

5.4.1. Dataset

This section describes the raw calibration data and the derived data set that is later used to determine a new neural network based velocity calibration function. The data were provided by Srama (priv. comm.) and derived calibration functions are described in Srama [2000] and Srama [2009]. All obtained data were parsed and stored in a proper SQLite database for further analysis. A current list of the database structure and the database itself are stored on the *idust* server.

Calibration Data

Srama [2000] and Srama [2009] provide miscellaneous calibration functions for CDA to determine the charge, mass or velocity of an impacting particle. Raw data from several CDA channels lead to different channel-depending functions that are listed in the chapters 3.4.1 and 3.5.1 (for IID, respectively CAT) in the first reference and in 3.3.3 in the second reference, respectively.

The velocity was determined with the dust accelerator at the Max Planck Institute for Nuclear Physics in Heidelberg (Germany) using differently sized and composed particles [Srama, 2000, 2009]. A complete list of the dataset was provided by Srama (priv. comm.) containing over 4500 calibration measurements for both targets as well as wall impacts. 2550, 1659 and 487 data points are available for CAT, IID and Wall impact calibrations. In general, 4 different impacting materials have been used, namely Iron, Aluminium, Carbon and Latex spheres. A

few experiments were done with additions or coatings, like Sodium in Carbon dust grains or Polypyrrol / Polystyrol, Polyaniline or Silicon coated Latex. Figure 5.52 shows a histogram of dust particle shots on both targets and the wall differentiated by the used material. The CAT was calibrated with all materials. Most measurements were performed with iron grains. Furthermore, CAT is the only target that was calibrated with Aluminium dust. IID's calibration consists mostly of iron shots, approximately 300 Latex shots and a few Carbon grains, where the Walls were calibrated approximately equally with iron and Latex. For both detectors, shots were performed on center parts of the targets as well as on the edges. The coordinates of the impact location are also stored in the calibration data set. However, since no coordinates can be derived of in-situ detections, all calibration data will be used for the velocity function computation.

Furthermore, for CAT impacts, the number of mass spectra peaks was determined and saved, too. Figure 5.53 shows the number of determined mass peaks. For a proper CAT calibration (mass & velocity) only data should be used with at least 2 mass peaks. 1554 measurements fulfil this requirement and contain detections with up to 20 peaks.

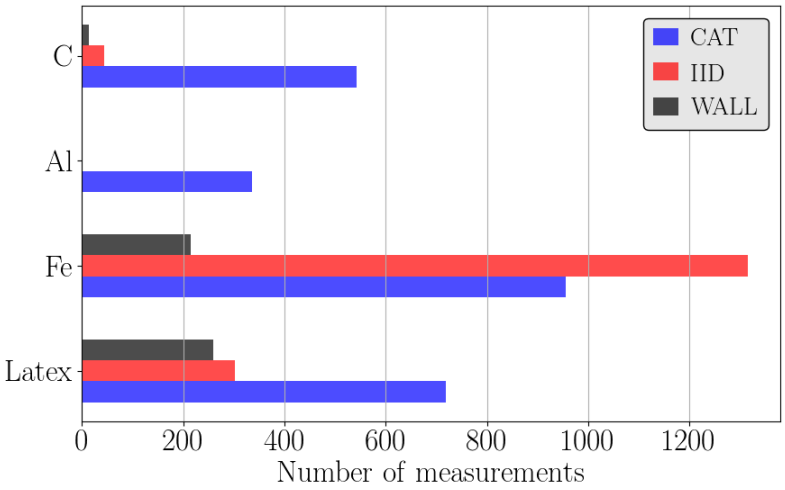


Figure 5.52.: Cosmic-Dust-Analyzer calibration materials. For each target (Chemical Analyzer Target (CAT), Impact Ionisation Detector (IID) and Wall) the number of used calibration shots in the dust accelerator is shown. The Y-axis differentiates the dataset by the used material.

Due to wrong or faulty measurements some rise times from different CDA channels and / or charges determined by the PSU[8] have a null value. Thus, for the velocity calibration, the dataset needs to be cleaned. Outliers need to be excluded from the dataset, too. For example, figure 5.54 shows the distribution of the QI rise times for CAT impacts. The number of events is shown logarithmically on the Y-axis. The main rise time distribution ranges between a few microseconds and up to 25 μs. Three additional data points are present beyond a rise time of 60 μs. Due to the insignificant number of large rise time measurements and the gap between the main distribution these 3 data points are neglected.

[8]The PSU is a crucial part of the dust accelerator that allows one to filter incoming dust particles by charge, speed and mass.

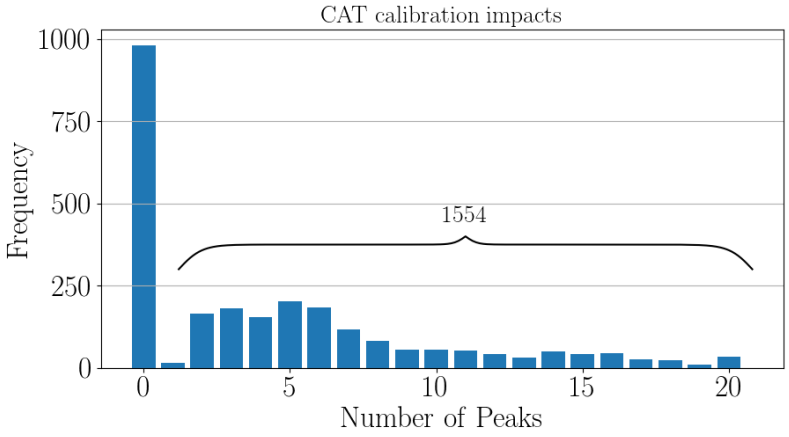


Figure 5.53.: Chemical Analyzer Target detections vs. the number of determined mass spectra peaks (by the Cosmic-Dust-Analyzer electronics). Only impacts with at least 2 mass peaks can be used for a proper Chemical Analyzer Target (CAT) calibration (mass & velocity). In total, 1554 data points fulfil this requirement.

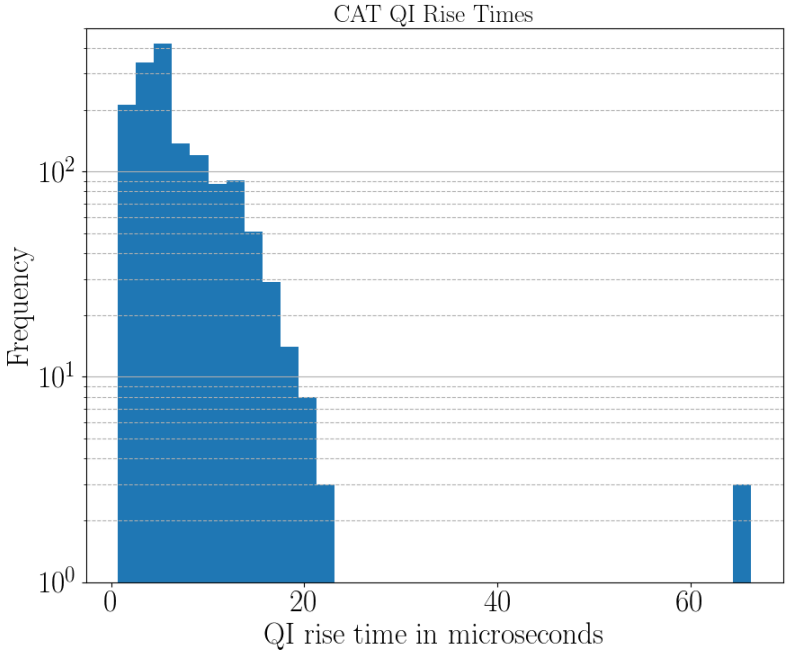


Figure 5.54.: Distribution of the QI rise times for the Chemical Analyzer Target (CAT) impacts (number of peaks at least 2).

The velocity calibration is only be valid for a certain mass range (and assuming spherical particles a correspond radius respectively). The PSU only determines the velocity v and charge q of the particle. A particle's mass m can be determined via the following equations. Equation 5.27 is the kinetic energy of a dust particle and equation 5.28 gives the energy in an electrostatic field. U is the voltage of the accelerator that is mostly set to 2 MV. Both formulas lead to the mass equation 5.29.

$$E_{\text{kin}} = \frac{1}{2} \cdot m \cdot v^2 \tag{5.27}$$

$$E_{\text{el}} = q \cdot U \tag{5.28}$$

$$m = \frac{2qU}{v^2} \tag{5.29}$$

Figure 5.55 shows the parameter domain of the calibration. The velocity of the particles is plotted vs. their charge. Both parameters are determined by the PSU and are given in km/s and fC, respectively. Each subplot shows the domains for CAT, IID and wall impacts and the different colors indicate different materials, namely Carbon, Aluminium, Iron and Latex (primary component; some dust grains were coated or mixed with additions). The charge ranges from around 1 fC to 100 fC for both targets. CAT however was calibrated with more low charged aluminium particles.

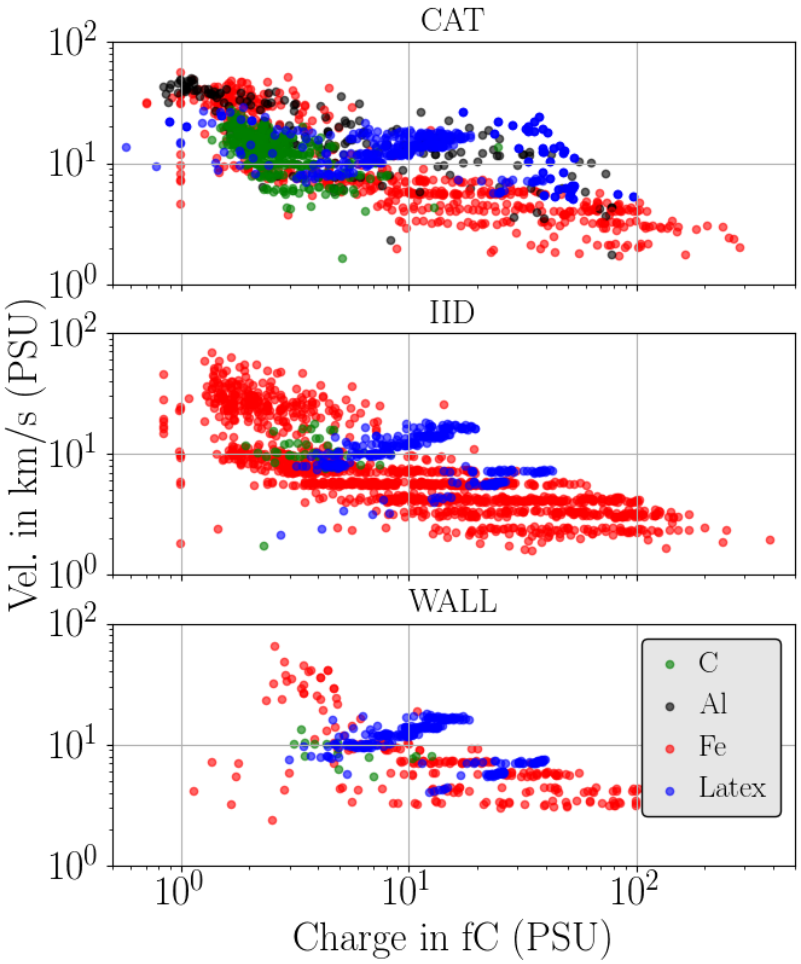


Figure 5.55.: Scatter plot of the calibration particles' velocity / charge parameter space. Both axes are shown logarithmically and the colors indicate different materials. From top to bottom: subplot for each target, namely CAT, IID and Wall, respectively.

Considering equation 5.29 the velocity and charge parameter range leads to a mass distribution that is shown in figure 5.56. The different colors indicate the calibration distribution for the two targets and the wall. The masses are given in picogram and range approximately from 10^{-3} to 10^{4} pg for all impact targets. Considering the density of each material and assuming a spherical dust particle shape CDA is calibrated for sub-micron to micron-sized particles. However, most particles within the Saturnian system consist of water ice. Different material properties and densities with varying porosities might lead to different CDA channel responses. Thus, the derived calibration functions are only valid within this certain mass range.

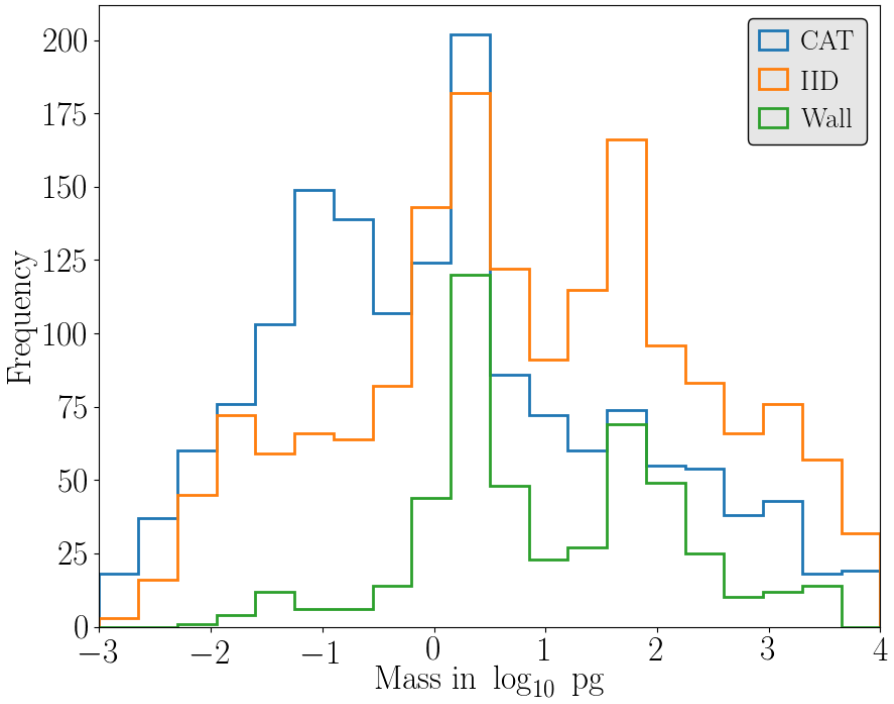


Figure 5.56.: Mass distributions of the used calibration particles for each target. The distribution is shown in $\log_{10}$ picogram.

More sophisticated velocity calibration functions should also consider different mass scales. However, due to the mass calibration bias of the dust accelerator this additional differentiation might not appear suitable. This work focuses mainly on a velocity (v) function depending on the rise time t_r and charge q of the individual channels $v\,(t_r, q)$. However one should consider the velocity – mass bias as shown in figure 5.57. Here, the velocity in km/s is shown vs. the determined mass in $\log_{10}$ picogram. Each color represents the three different targets. The velocity is anti-proportional to the mass. Consequently, for a proper particle property determination the following steps could apply (neglecting that no calibration measurements were performed with water ice (solid and / or fluffy)):

1. Determine the mass of the particle (see derived equations by Srama [2000])
2. Determine velocity of the particle (either with the already derived calibration functions or the neural network regressions shown in this work)

3. Filter particles out that are not within the calibration range shown in figure 5.57

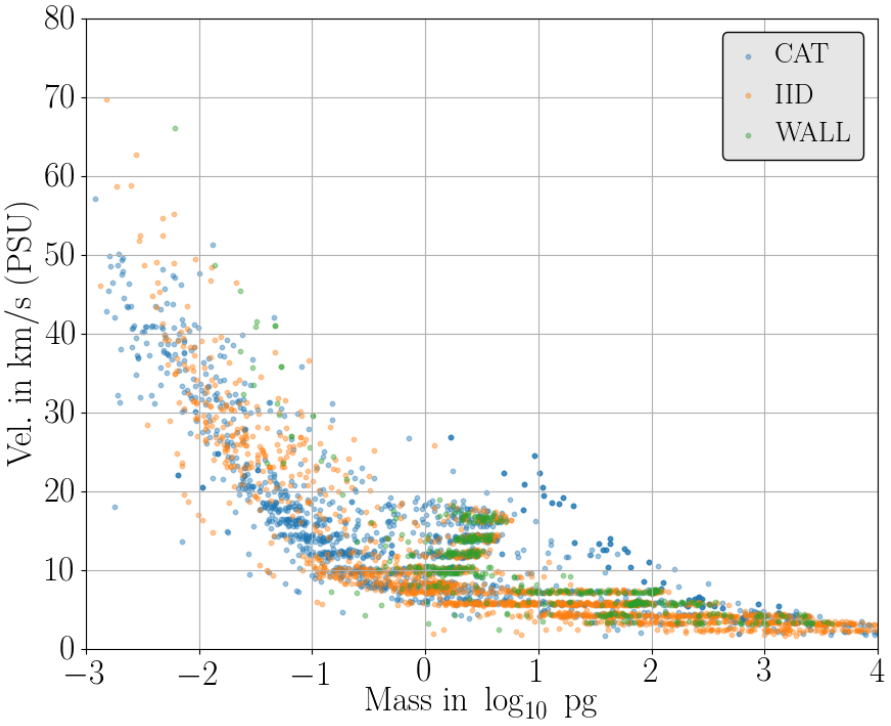


Figure 5.57.: Scatter plot of the calibration particles, where the particles' velocity is shown vs. their masses. Three different colors represent the different targets.

Velocity Calibration Data

This works shows the neural network based regression models for 5 different impact cases, namely CAT QC and QI, IID QT and QI and Wall QI impacts. After filtering wrongly determined rise times and / or charges the following number of calibration points is used for the regression models. Table 5.5 lists these numbers. Most calibration points are present for IID impacts with QI signals. At least 479 calibration points are available for Wall impacts with suitable QI data.

Table 5.5.: Number of velocity calibration data points for 5 different impact scenarios.

Target	Channel	Calibration Points
CAT	QC	1535
	QI	1510
IID	QT	1547
	QI	1642
Wall	QI	479

Exemplarily, the calibration data for IID impacts (QI channel) are shown in figure 5.58. The velocity of the dust grains determined by the PSU is plotted vs. the rise time and charge

determined by QI. Rise times and the charge are given in μs and $\log_{10}$ fC, respectively. The velocity is given in km/s. For better readability, the color code represents the velocity, too. A color-bar at the bottom of the figure shows the according scale.

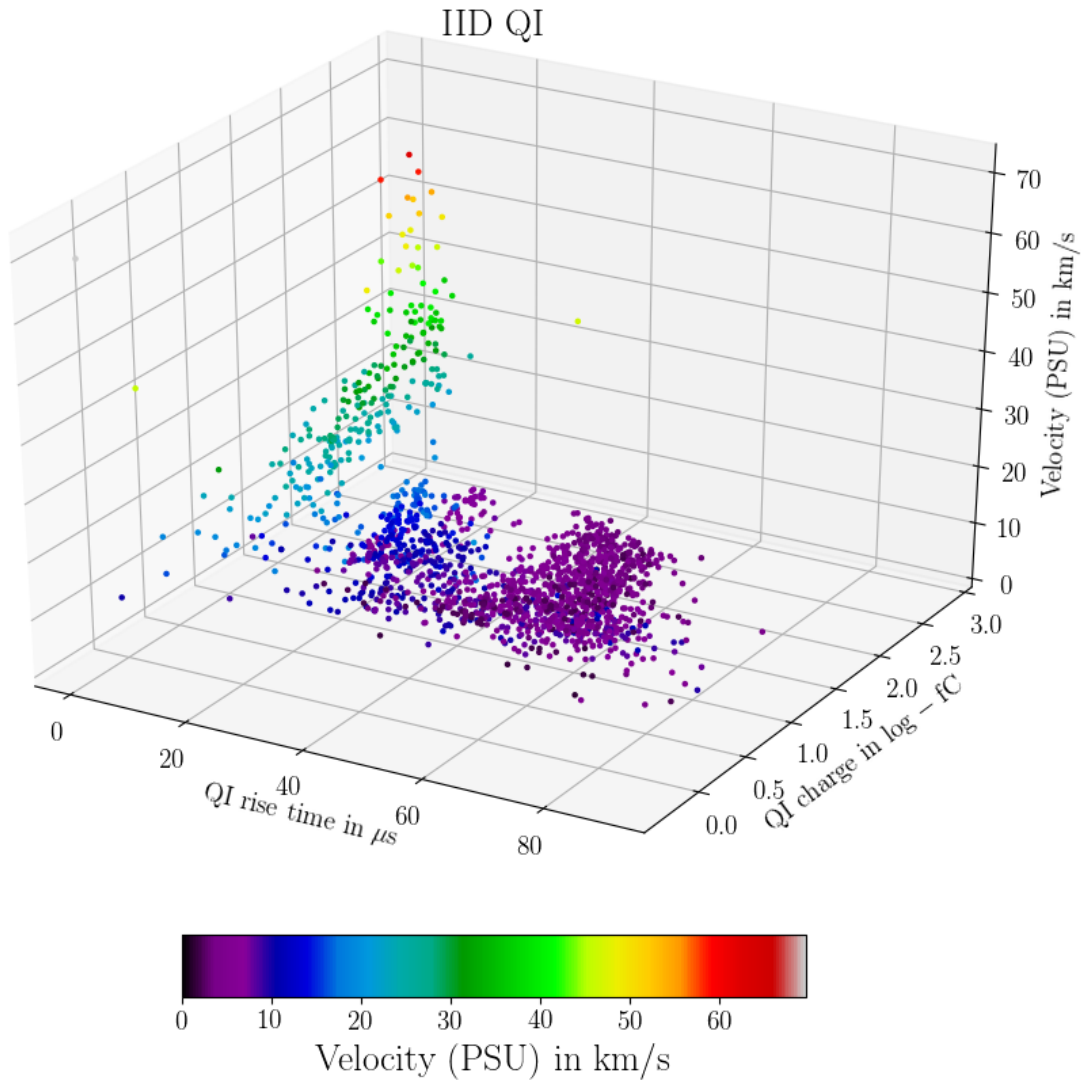


Figure 5.58.: Impact Ionisation Detector (IID) calibration data (QI). The particles' velocity in km/s is shown vs. the rise time in μs and charge in fC. For better readability, the velocity is also color-code.

The data set is approximately divided into two domains. Fast particles correspond to a short rise time. Additionally, the determined velocity increases for increasing charges. Within 3 orders of magnitude for the charge, the velocity increases from around 20 km/s up to 70 km/s. A velocity calibration function that neglects this effect might over- or underestimate the according velocity. The second domain represents the calibration points for higher rise times larger than around 10 μs. Here, a possible calibration plane has a less steeper gradient and shows only slight charge depending effects.

The remaining figures for CAT QC and QI, as well as IID QT and Wall QI are shown in the appendix C.5.2 (merged with the resulting regression planes).

5.4.2. Neural Network Regression Models

This section describes the architecture of the velocity regression NNs, where a channel's rise time and amplitude are the corresponding input parameters. A detailed view on the IID QI results and further regression plane properties is provided. All other models are shown in detail in the appendix C.5.1.

First, the miscellaneous architectures and the training procedure are shown. Then, the results are discussed and how the models are generated for the mean velocity and corresponding error function.

Neural Network Architecture

Since there is no default programmatic "blue-print" to develop a NN for a regression problem, in this work a few NN architectures are developed and tested. Simple solutions, as well as more sophisticated, adaptive networks are used to determine the best model for the calibration data. For the regression the amplitude and rise time of the different channels are available as well as the corresponding velocity determined by the PSU. For every channel individual models are developed. The number of input neurons, and the number of hidden neurons and layers need to be optimised. Only the output layer's design is pre-defined, since only one velocity value per amplitude and rise time is computed.

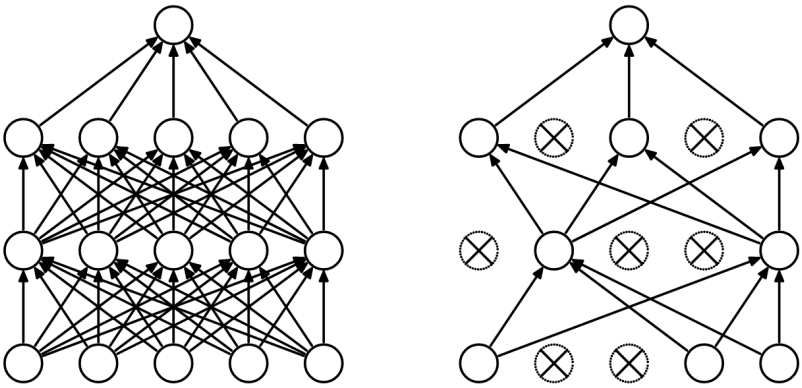
In total 6 (feedforwards) models are defined with Rectifier Linear Unit activation functions for each neuron. The regression problem is non-linear, however several Recitfier Linear Unit neurons and layers can describe any non-linear function. One model called *simple* is a 3 layer model with only a few neurons. 4 layers are implemented in the model *default*, where also different dropout rates are applied. Srivastava et al. [2014] introduced miscellaneous dropout models. A new hyperparameter sets how many neurons shall be removed with a certain algorithm. By removing neurons the NN does not tend to overfit. By applying miscellaneous dropout rates for a model, the rate resulting in the best score is the preferred solution. Figure 5.59 shows schematically the dropout method on a 2 hidden layer NN. It is shown (subfigure b)) that neurons that are dropped (marked with an X) do not contribute to the model. The disadvantage of the dropout NN is the higher training time. Compared with an identical model with no dropout rate, the training time is 2 - 3 times higher.

Then, a third model called *deep* is developed, also with an applied dropout rate. The large number of hidden layers is set, but is way too deep and complex for a rather simple regression problem. It is set for testing purposes to determine, how severely such models tend to overfit. In all cases that are later shown, *deep* models do not perform well.

Finally, the *adaptive* model, with three different dropout models is considered. This model is based on ideas from Huang [2003], who determined a "rule-of-thumb" for 2 hidden layers feedforward NN. Here, the case is even more simplified by using only 1 hidden layer in a 3 layer structure. The input layer is also set adaptively. A large number of neurons is used that are

sensitive for different pairs of rise times and amplitudes.

With all models set, the training of the models is performed.



(a) Standard Neural Net (b) After applying dropout.

Figure 5.59.: Schematic visualisation of a Neural Network with an applied dropout rate. Left a): Original neural network with 2 hidden layers and 16 neurons in total. Right b): Neurons with an X do not contribute to the network any more (in total 7 neurons). From Srivastava et al. [2014].

Training and Verification

The training of the NN is rather straightforward. Within a larger iterative for-loop all models are selected and tested with all available parameters and options. For a high performance computation, the models were trained on multiple central processing units.

All models, with different dropout rates (if available), batchsizes, etc. are trained and the corresponding scores are determined and stored. After collecting all training results, a few randomly chosen models are visually inspected, to determine, whether the regression models appear feasible. Then, the collected data are analysed.

Results

As described before several Neural Networks are trained per data set. Table 5.6 shows the results for the IID QI dataset. Each row lists the best mean (μ) score (normalised) with its corresponding (σ) error for each model. The models, as described in the first section, are listed in the first column. The second, third and fourth columns show the dropout rate, batchsize and number of epochs for the best fitting model. Note: The model *Simple* does not have a dropout rate and the *Adaptive* model with Gaussian-Noise requires a noise factor σ. For the IID QI data set the *Default* model with a dropout rate of 0.1, a batchsize of 30 and 8 epochs provides the best regression model. The results for the remaining data set are shown in appendix C.5.1. An *Adaptive* model with a Gaussian-Dropout performs better for all other data sets. The dropout rate for all models is rather low but for the Wall impacts (QI data). For the wall data, two *Adaptive* models provide the same score (green and orange colored). However, the Gaussian-Dropout model shows a lower error and is used as a regression model.

Table 5.6.: Neural Network regression results for the Impact Ionisation Detector QI data. The first and second column list the model and dropout rate. However the last row shows the results for the Gaussian Noise layers and gives the σ noise. Column 3 and 4 show the batchsize and epoch for the best regression model. The according score and standard deviation of the mean score is shown in the last two columns. Both parameters are given in the scaled domain.

Model	Dropout	Batchsize	Epoch	μ Score (scaled)	σ Score (scaled)
Simple	—	15	8	0.514	0.089
Default	0.1	30	8	0.389	0.063
Deep	0.0	20	32	0.44	0.075
Adaptive (Dropout)	0.0	25	32	0.427	0.086
Adaptive (GDropout)	0.0	30	64	0.411	0.053
	σ Noise				
Adaptive (GNoise)	0.1	15	16	0.403	0.064

The complete mean score domain for the *Default* model can be seen in figure 5.60. The mean score is plotted vs. the batchsize and the number of chosen epochs. For a better readability the dots are linked with a height depending color plane. Lower scores, and consequently a better regression model is indicated with a brighter color and vis versa. The best fitting parameter as shown in table 5.6 is shown as a red triangle. The resulting domains for the other data set regression models are displayed in the corresponding appendix.

A small batchsize and a large number of epochs result in a bad score that improves asymptotically for higher batchsizes and smaller epochs. The regression model for the IID QI data converges perfectly for higher numbers and leads to the best scoring parameter. A batchsize of 30 and 8 epochs is the local minimum in the computed domain. The other regression models, as shown in the appendix, do not converge as smoothly as the model described here. Only the CAT QI model converges smoothly while other regression models show local (strong) fluctuations and worse scores.

Model generation

Based on the best scores, the final neural network regression models are computed for each dataset. Due to the lack of large rise time calibration points the final model for the CAT QC data are limited to 5 μs.

Figure 5.61 shows the resulting calibration plane for the IID QI data set. This 3D plot is similar to the data scatter plot shown in figure 5.58. The velocity of the particles, determined by the PSU is shown vs. the rise time and charge determined by the QI channel, respectively. Both parameters are given in μs and $\log_{10}$ fC. Again, the scatter points represent the calibration points and the color indicates the velocity. A color-bar that encodes the velocity is shown at

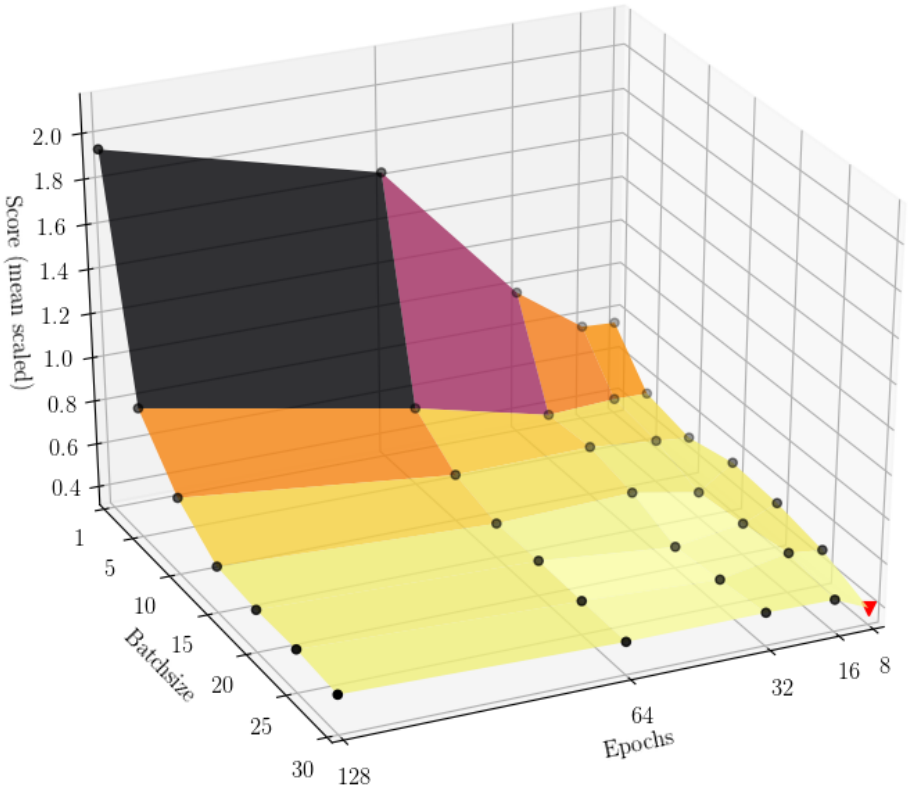


Figure 5.60.: Score domain for the Impact Ionisation Detector QI regression model (*Default* model with a 0.1 dropout rate). The scaled mean score is plotted vs. the used batchsize and number of epochs. For better readability, the dots that represent the scores are connected with height depending color planes, where brighter colors indicate a better score and vis versa. The best solution is indicated with a red triangle.

the bottom of the plot. The plane is the neural network regression model. Similarly to the alright existing calibration function, the neural network determined the decreasing velocity for higher rise times. The slope of the function changes and has its highest values for very short rise times. Additionally, the regression model describes the increasing velocity for short rise times and increasing charges. An increasing charge of two order of magnitudes leads to a velocity increase of approximately 2 times (from around 20 to 40 km/s). This slope vanishes for higher rise times and reverses slightly. However, this reverse is probably a slight overfitting effect of the regression model. To determine the stability of this solution, the final model generation is repeated several times. All solutions with randomised initial states lead to the same solutions.

All other solutions are shown in the corresponding appendix of this work. The regression model for the IID QT data set is also a stable solution with a similar calibration plane shape as for the QI data. For the QT data, the reverse gradient for higher rise times is also present.

Unfortunately CAT data are less stable. QC data lead to a similar solution as for both IID data sets, however the QI data appears overfitted. Furthermore, the function decreases significantly for rise times after approximately 2.5 μs (local maxima). This is probably due to

overfitting and / or worse data quality. For large charges and rise times the calibration plane predicts negative impact velocities, due to the lack of data points.

Surprisingly, also the solutions for wall QI impacts are stable and show again a similar shape as the other calibration planes. However, the number of data points with short rise times and large charges is limited. Thus the resulting velocity values in this domain need to be seen critically.

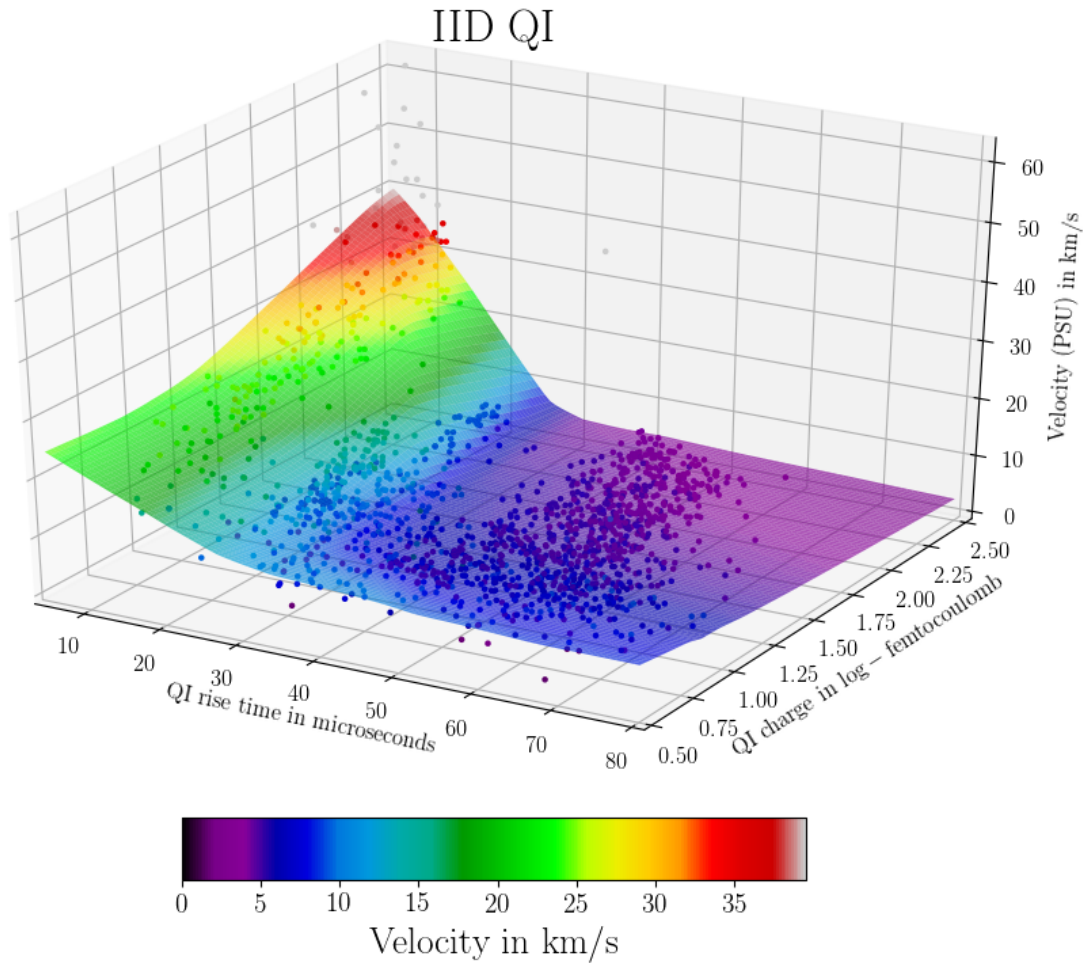


Figure 5.61.: Impact Ionisation Detector (IID) QI calibration data with the computed neural network regression model (plane). The velocity in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

All models cannot be displayed analytically, but are stored on a computer as an object (data format is HDF5) for the library Keras. Extracting the weight matrices, activation functions and the normalization method, one can adapt the results to other programming languages. Furthermore, compared with the calibration functions by R. Srama, the mathematical domain, where the functions are valid, is way smaller. Input amplitude and rise time are defined in a strict range, as can be seen in the following figure. Srama [2009] on the other hand did not set a valid amplitude range.

Error Determination

In the previous section, neural network regression models have been shown for the mean velocity function, depending on the channel's rise time and determined charge. This section shows the corresponding error determination for each regression model.

For each regression model, the KERAS library provides a corresponding score that can be interpreted as the average deviation between the fitted plane and the individual data points. Figure 5.62 shows the distribution of the ratio of the velocity determined by the fit vs. the actual data points (IID QI). The black solid and dashed lines indicate the median and corresponding quartiles, respectively. In average, the neural network regression fit describes the data points well. The data points scatter around the regression plane that leads to the shown non-symmetric distribution and one can derive that the regression model overestimates partly the velocity.

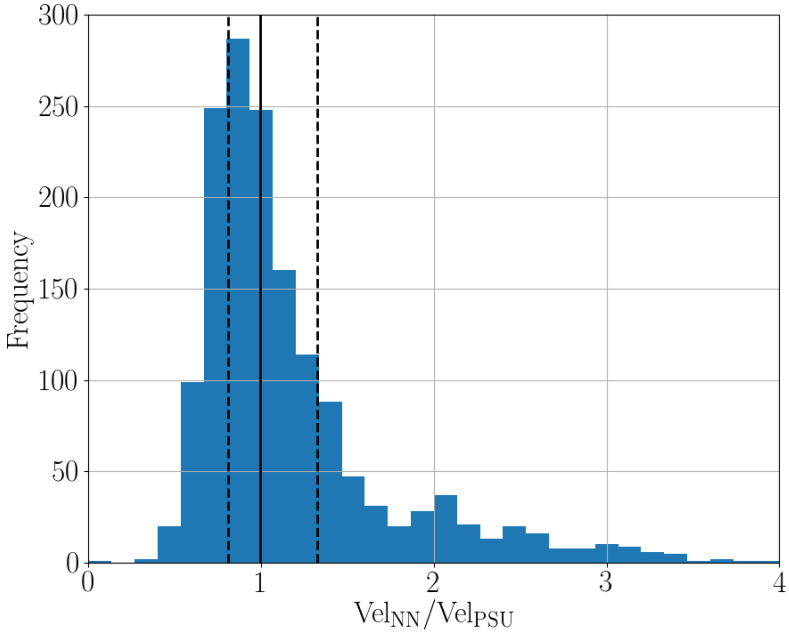


Figure 5.62.: Distribution of the deviations between the regression fit (Neural Network (NN)) and the actual measurements (Impact Ionisation Detector (IID) QI data) from the Particle Selection Unit (PSU). The black solid and dashed lines indicate the median and corresponding quartiles of the distribution, respectively.

The error of the velocity determination however is not a fix value but depends on the rise time and charge. The regression models as shown previously and in the appendix show larger deviations for shorter rise times than for larger ones. Thus, additional models need to be computed for the corresponding error.

To determine a sophisticated error model, first, the absolute deviation between the data points and the regression plane is determined. The deviations are given in km/s and the resulting data for IID QI impacts are shown in figure 5.63. Here, the deviation is plotted vs. the rise time and charge. A color code represents the velocity for better readability. The shown plane is the resulting model for the error. A robust scaling is applied. To prevent a too

complex and overfitting solution, especially for highly scattered data for short rise times, the simple neural network model is used (for all data sets).

All error models have a similar shape, probably due to the same scatter point distribution. Short rise times correspond to a larger deviation while the deviations appear rather constant for larger rise times. The resulting error models consist of two parts: An approximate linearly decreasing plane for short rise times and a constant plane for the large ones. For IID QI data the error has a maximum of around 9 km/s for very short rise times and decreases down to 0.5 to 1 km/s.

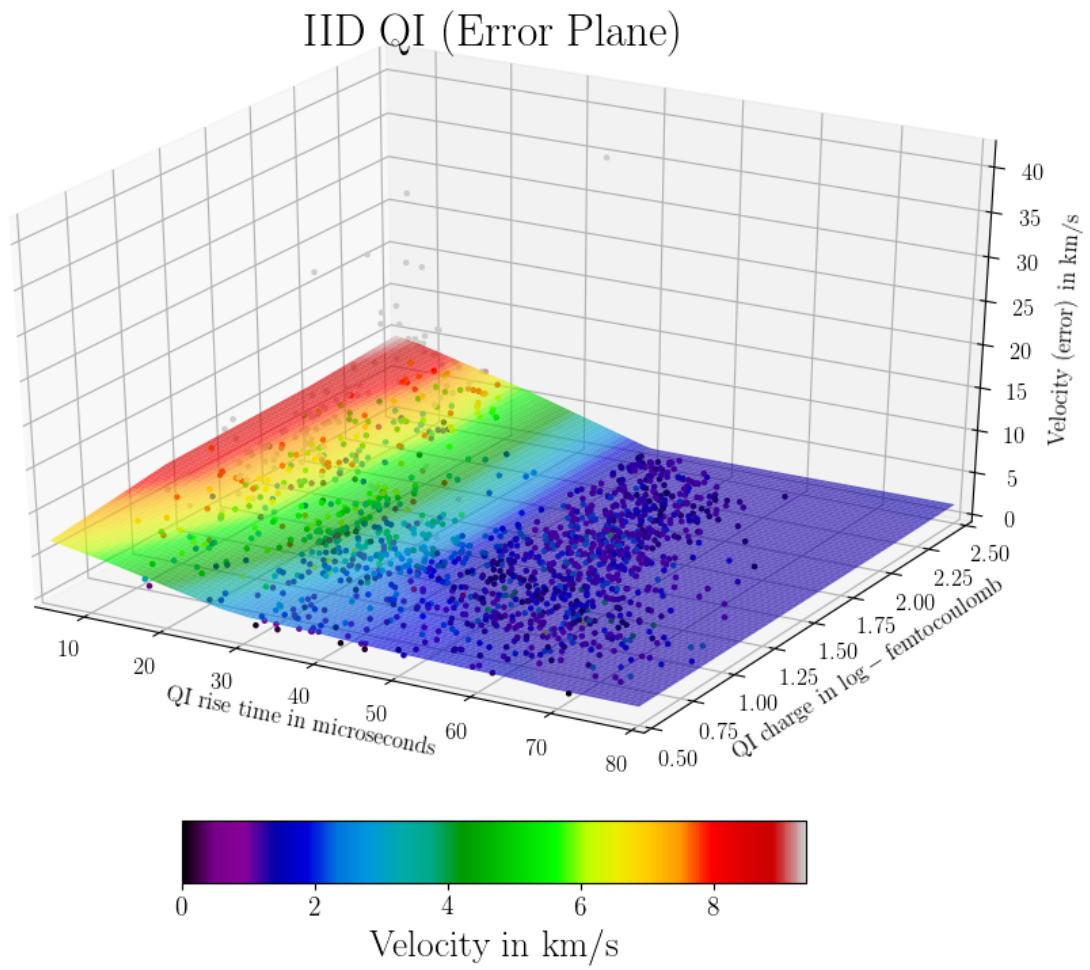


Figure 5.63.: Impact Ionisation Detector (IID) QI deviation data with the computed error model (plane). The velocity deviation in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

5.4.3. Test & Verification

This section describes the performance of the determined models (miscellaneous comparisons with e.g., Srama et al. [2004] function).

Comparison regression model – PSU

First, the performance of the regression models is determined. To quantify the performance, the velocity determined by the calibration functions is compared with the corresponding PSU data. The resulting data sets of each target and channel are shown in figure 5.64 to 5.66. The dots represent the PSU vs. regression data and the black line shows the theoretically perfect calibration line (no deviations between calibration function and measurement data).

Figure 5.64 shows the results for the IID QT and QI data, respectively. The data comparison scatters well around the ideal calibration line. For lower speeds, the calibration function returns systematically too large velocity values. E.g., a median value of around 4 km/s is determined although the corresponding PSU measurement is approximately 1.5 km/s. However, most impacts on CDA are higher in the Saturnian system. Thus, a bias would affect only a small fraction of the dataset. High speed impacts, in the Saturnian system e.g., interplanetary or interstellar particles, appear to be well calibrated for both channels. With a neural network regression model an improvement of low speed impacts is not possible. Similar issues were also described by Srama [2009, chap. 3.4 p. 54]

A good calibration can also be seen for CAT QC data, where the low speed domain is not biased. Slightly worse is the QI data set. However, as already described, the regression model is overfitted for this dataset, and not suitable for a proper data analysis.

Figure 5.66 shows finally the comparison of the wall calibration function with the corresponding PSU data. Between 7 km/s and 11 km/s the calibration suits well. For lower speeds, the regression model shows a similar trend as the functions determined for both IID channels. For higher velocities, no proper bias can be determined due to the lack of data.

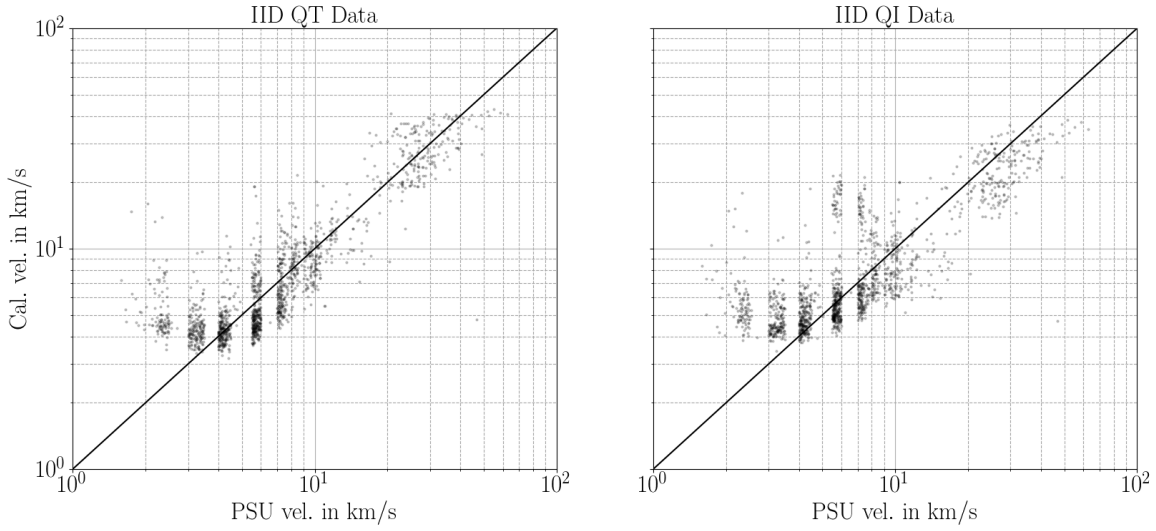


Figure 5.64.: Comparison of the regression model data vs. the corresponding measurements data determined by the Particle Selection Unit (PSU) (Impact Ionisation Detector (IID); left sub-figure: QT, right sub-figure: QI). The velocities are given in km/s and are plotted logarithmically. A black solid line indicates a theoretical perfect calibration line, where the velocity determined by the function describes the actual data perfectly.

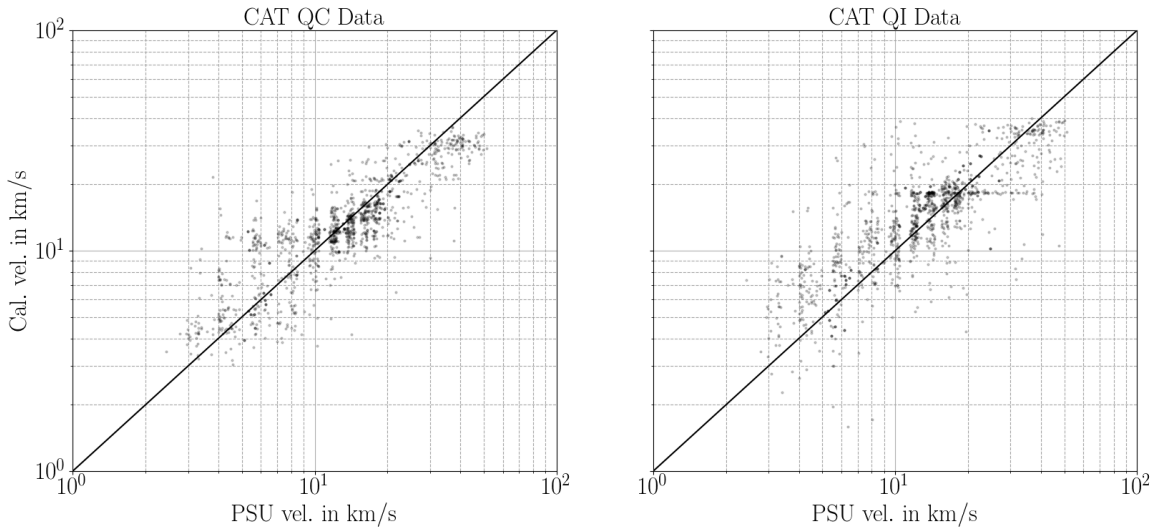


Figure 5.65.: Comparison of the regression model data vs. the corresponding measurements data determined by the Particle Selection Unit (PSU) (Chemical Analyzer Target (CAT); left sub-figure: QC, right sub-figure: QI). The velocities are given in km/s and are plotted logarithmically. A black solid line indicates a theoretical perfect calibration line, where the velocity determined by the function describes the actual data perfectly.

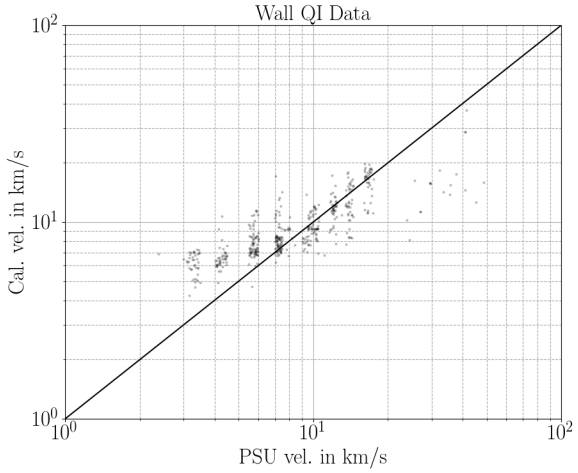


Figure 5.66.: Comparison of the regression model data vs. the corresponding measurements data determined by the Particle Selection Unit (PSU) (Wall; QI). The velocities are given in km/s and are plotted logarithmically. A black solid line indicates a theoretical perfect calibration line, where the velocity determined by the function describes the actual data perfectly.

Inter-Channel deviations

In this section both channel models (large and small target) are compared. As shown above, the regression models fit the data quite well. Slight deviations are present in the IID model for low velocities, and the QI regression model for the small target is not as suitable as the QC dataset. Here, the QT - QI and QC - QI results are compared for IID and CAT respectively.

Figure 5.67 shows this inter-channel comparison for IID and CAT in the left and right sub-figure, respectively. The QI velocity data are plotted vs. the corresponding QT and QC data. A black solid line is a linear function and shows the same determined velocities. The models

for both IID velocity functions provide approximately the same results. Low velocity as well as high speed particle impacts are described similarly by both functions. A higher deviation or data scattering appears between 6 km/s and 10.5 km/s, where some velocities determined by the QI function are higher than the ones determined by the QT model. Here, most velocities however are slightly below the solid line.

The right figure shows a similar plot but for QC velocities on the X-axis instead of the QT data as shown for the IID. Due to the overfitted QI function the variation and deviations are quite large. The IID data show a scattered, but rather dense distribution around the ideal solid black function. For CAT the scattering is especially severe for low velocities but improves for high impact speeds. Whether these biases can be improved by rise time or amplitude constrains needs to be determined and is shown in the following.

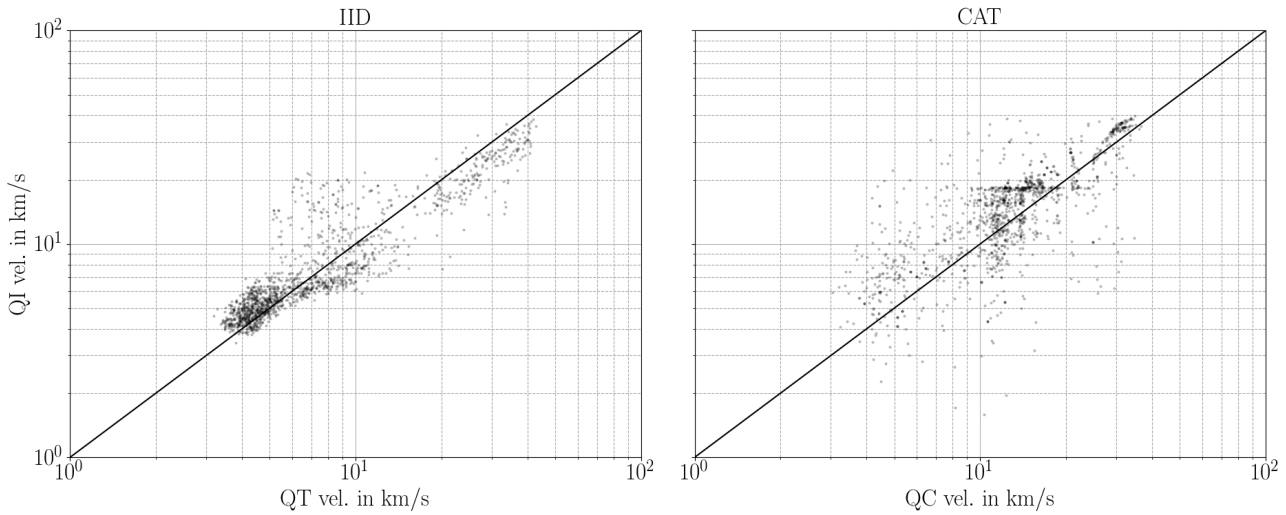


Figure 5.67.: Comparison of the velocities determined by the QT and QI data for the Impact Ionisation Detector (IID) (left sub-figure) and QC and QI data for the Chemical Analyzer Target (CAT) (right sub-figure).

CAT QI improvement The CAT data in figure 5.67 show a large scattering, since the QI regression model tends to overfit. Only high speed impacts larger than 25 km/s cluster denser around the ideal calibration line.

For the data analysis, testing and performance check, all data from all channels are stored in a dataframe. Thus, querying the data with certain constrains can be easily done. To determine the contrains of the denser high velocity sub-dataset, the amplitude and rise time domains need to be set accordingly. Figure 5.68 shows an improved version of the right sub-figure 5.67, where most largely scattered data are filtered out. Here, only data are considered with an amplitude of at least 100 fC and a rise time shorter than 10 μs. The regression model fits well for this impact domain where the resulting velocity is between 25 km/s and around 40 km/s.

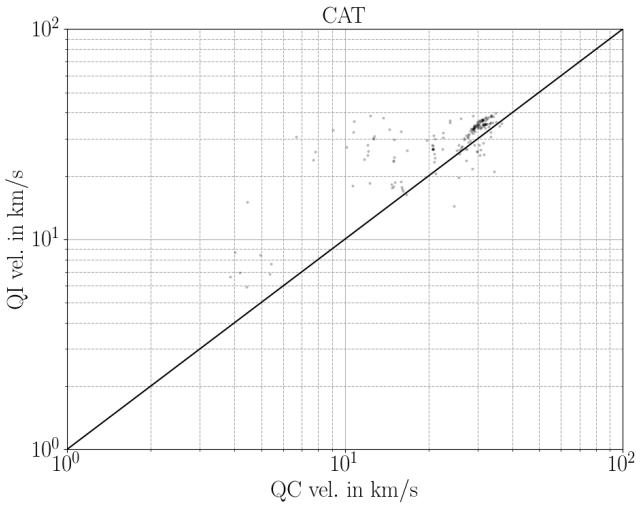


Figure 5.68.: Improved Chemical Analyzer Target (CAT) calibration data. This figure is based on the right sub-figure 5.67. Only data are displayed with an amplitude of at least 100 fC and a rise time less than 10 μs.

Comparison with existing calibration functions

The MC based program DOCC that computes the orbital elements of detected dust particles in the Saturnian system uses the IID QI and CAT QC velocity calibration function as already shown in the DOCC chapter. Srama [2009] computed also a velocity calibration function for QT signals (IID impacts), but due to insufficient data quality no functions for CAT QI and wall QI events. Consequently only three regression models can be directly compared with already existing calibration functions. Equation 5.30 gives the QT velocity function v, where t_r is the rise time in μs. Furthermore the function is only valid for amplitudes larger than 65 fC and a rise time between approximately 1 and 80 μs [Srama, 2009]:

$$v_{\mathrm{IID}}^{\mathrm{QT}}(t_r) = -240 + 361 \cdot t_r^{-0.117} + 0.338 \cdot t_r \tag{5.30}$$

The error factor of the function is 1.67. For $v_{\mathrm{IID}}^{\mathrm{QI}}$ and $v_{\mathrm{CAT}}^{\mathrm{QC}}$ the corresponding errors are 1.62 and 1.42, respectively. The developed DOCC program uses an outlier cleaned and scalable error. For this analysis the regression models are compared with the functions and errors provided by Srama [2009].

Figure 5.69 shows the computed mean velocity determined by the neural network regression models vs. the mean velocity determined by the calibration function by Srama [2009]. Three different colors represent the data for CAT QC, IID QT and QI, respectively. The velocities are given in km/s and a solid black line guides the reader's eye showing velocity values that have the same value determined by both functions. The velocity equations determined by Srama [2009] are only depending on the rise time. High velocities correspond to a short rise time. For high speed impacts one can see the discretisation of the the velocities, since these values correspond

to only a fraction or few μs (QC sampling rate: 6 MHz). For slow impact speeds both IID regression models overestimate the resulting velocities, compared to Srama [2009]. Beyond 6 km/s the resulting velocities from the neural networks are smaller than the values from the original calibration functions. For higher impact speeds the three neural network models provide slower velocities than the Srama [2009] functions. However, the variation between both models is not linear.

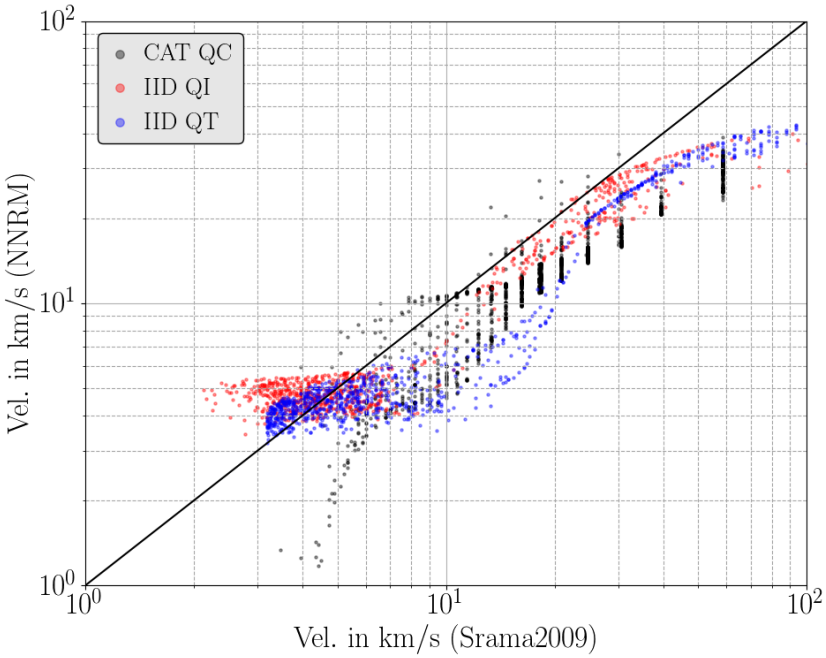


Figure 5.69.: Velocity determined by the neural network regression model (NNRM, Y-axis) vs. calibration by Srama [2009] (X-axis). The colors represent the results for the different calibration models. A black solid line indicates identical velocity values and can be used to guide the eye.

Determining the error of Srama's equation is straightforward. Based on the rise time, the function returns a corresponding mean velocity value. Applying the listed factors on the according velocities lead to the upper error limits. The relation between the mean and error velocity is shown in figure 5.70 for the CAT QC data and in figure 5.71 for IID QT and QI data, respectively. Here, the error is plotted in km/s vs. the mean velocity. Black solid lines indicate the resulting data range of Srama's work.

Smaller colored dots indicate the velocity depending error values based on the neural network regression models. The models cannot be computed in an analytical way as the equations from Srama [2009]. To compute a mean and corresponding error value the models need to be loaded and parsed, since they are stored digitally as objects. Furthermore the velocity depends on the rise time as well as on the amplitude determined by the channel. However, the 3D regression planes that have been shown and explained indicate basically 2 trends: shorter rise times lead to higher velocities, and higher amplitudes lead to higher velocities too (significantly only in the short rise time domain). In the shown figures, the amplitude is color coded in $\log_{10}$ fC.

The resulting regression model improves the velocity determination error significantly. The

CAT QC data show (figure 5.70) that the improvement ranges up to a factor of 4 to 5. Further, the neural network model provides worse results for approximately 10 km/s particles that have a small charge (a red circle indicates these data points in the figure). Apparently the regression model did not provide a sufficient fit in the "transient" region between the short and long rise time data points. In general however the regression model improves the error computation. Figure 5.71 shows that the IID velocity error models provide smaller errors as well.

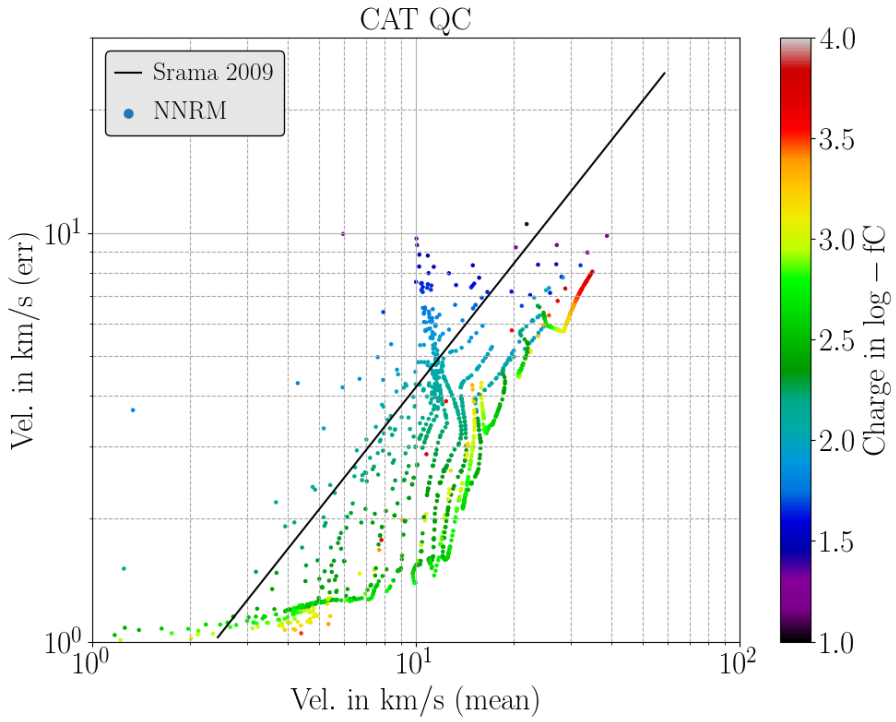


Figure 5.70.: Error of the velocity determination determined by the neural network vs. the calibration function by Srama [2009] (for the Chemical Analyzer Target (CAT) QC). Calibration data are used for comparison. The black solid line represents the data based on Srama's function. The dots represent the error based on the neural network, where the color indicated the charge of the particle. A red circle indicates slow particles with a relatively small charge, where the regression models overestimates the velocity compared to Srama [2009].

Performance

All final NNRMs are stored as an object for later computational purposes. Since using the models requires some programming overhead a user-friendly shell-based program is build to compute the velocities easily. Since the PYTHON based program uses several of these objects a performance check is required to determine the feasibility of this computational method. The computer that is used for this performance check has the following specifications:

- Core i7 4702MQ (4x 2,2 GHz)
- 8 GB DDR3 PC3-12800

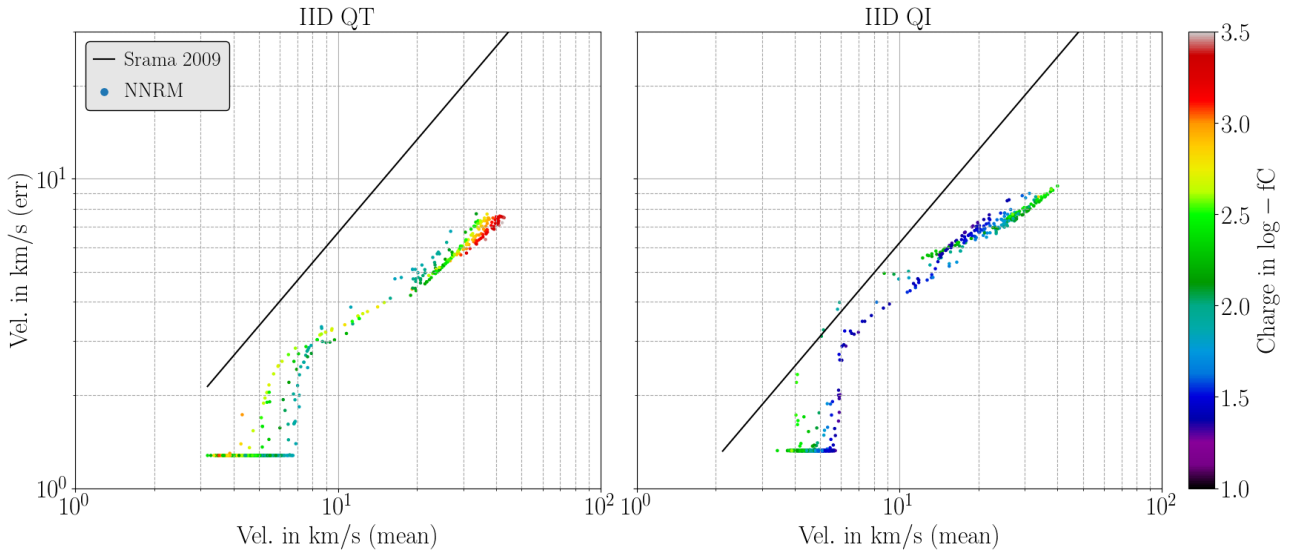


Figure 5.71.: Error of the velocity determination determined by the neural network vs. the calibration function by Srama [2009] (for the Impact Ionisation Detector (IID) QT and QI). Calibration data are used for comparison. The black crosses represent the data based on Srama's function. The black line represents the error based on the neural network, where the color indicated the charge of the particle.

Running a velocity computation requires several module and object loadings. Besides the NNRM object that is stored in an HDF5 format, different scalers in the *pickle* format need to be loaded, too (amplitude and rise time scaler). Loading the modules, classes and objects takes between 8 and 12 seconds. Loading the TENSORFLOW backend for the KERAS library is the most time consuming bottleneck (around 7 seconds). Loading the objects into the memory takes in total around 3 seconds. However once loaded, computations can be performed without re-loading the modules.

For the performance check, the scalers, definition file and NNRM of the IID QI channel are loaded. The number of computed velocities is varied from 1 amplitude and rise time value pair to 1,000 pairs. The definition range of the channel regression model is, where t_{rise} and q_{amp} is the rise time and amplitude, respectively: $5.0\ \mu\mathrm{s} < t_{\mathrm{rise}} < 80\ \mu\mathrm{s}$ and $10^{0.5}\ \mathrm{fC} < q_{\mathrm{amp}} < 10^{2.5}\ \mathrm{fC}$. The variation of the values is uniformly distributed within the definition range[9].

For each pair 10 individual computations are performed to determine a sophisticated statistical overview of the algorithm's performance. For each computation pair a box plot is calculated as shown in figure 5.72, where the computation time is plotted vs. the number of input pairs. Note that both axes are logarithmically scaled. The red solid line that crosses all boxes indicates a linear computation time increase. Thus, increasing the number of input pairs N results in a computation time increase of $\mathcal{O}(N)$. However, the computation time (after loading all kernels and modules) is very short. Computing 1,000 input pairs takes in average around 1.4 second to finish.

[9] All definition domains of all functions are stored in a so called definition file that is part of the developed library CDACAL (see conclusion).

The fast computation, even for larger input pairs, and the simple linear correlation between the number of pairs and the computation times, shows that the NNRM algorithm is fast enough for further applications.

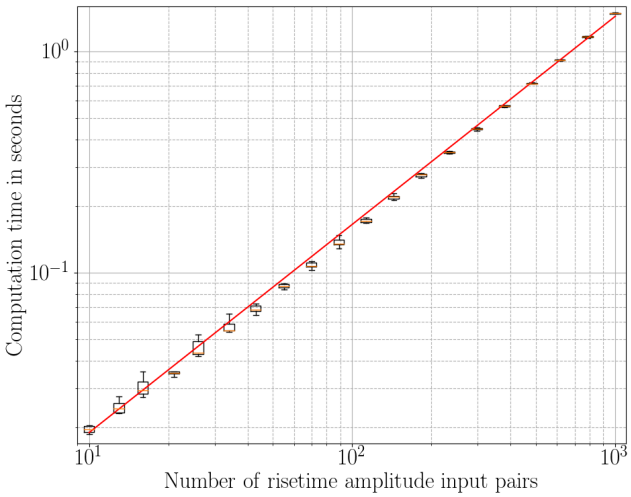


Figure 5.72.: Calculation time performance of the Impact Ionisation Detector (IID) QI Neural Network Regression Model (NNRM). The double-logarithmic boxplot shows the computation time for different input pair sizes (rise time and amplitude pairs). The red solid line crossing the boxes indicates a linear trend. Increasing the number of input pairs N leads to a linearly increasing computation time $\mathcal{O}(N)$.

5.4.4. Application on scientific data

The first version of the CDA orbit database computed the velocities of IID impacts with the rise time from the QI channel and Srama's calibration function from 2009 [Srama, 2009]. Furthermore, the corresponding science database query did not clearly consider the function's contraints. In this section, the results from the Srama calibration function are compared with the results from the NNRM on actual measurements from the Saturnian system. For the test application, measurements from the year 2008 are used, where a proper IID QI rise time and amplitude are determined. In total 60,427 particles are used.

Figure 5.73 shows the used particles' properties. The scatter plot shows the $\log_{10}$ amplitude and $\log_{10}$ rise time in fC and μs, respectively. Each dot represents a detected particle. The red and blue rectangles indicate the definition range of the Srama [2009] and NNRM functions, respectively. Note that the Srama [2009] function does not have an upper amplitude limit. The purple area is the overlapping definition range domain of the functions. The NNRM covers more particles with a smaller amplitude than the other calibration function. Srama [2009] however is valid for an infinite large amplitude value.

Considering the amplitude and rise time definition ranges for the 2009 function, 25,656 dust particle impact measurements can be used. This corresponds to 42.5 % of the chosen dataset. For the NNRM algorithm 55,804 measurements can be used; that is around 92.3 % of the dataset. Thus, the calibration function from Srama [2009] covers a factor of 2 less measurements

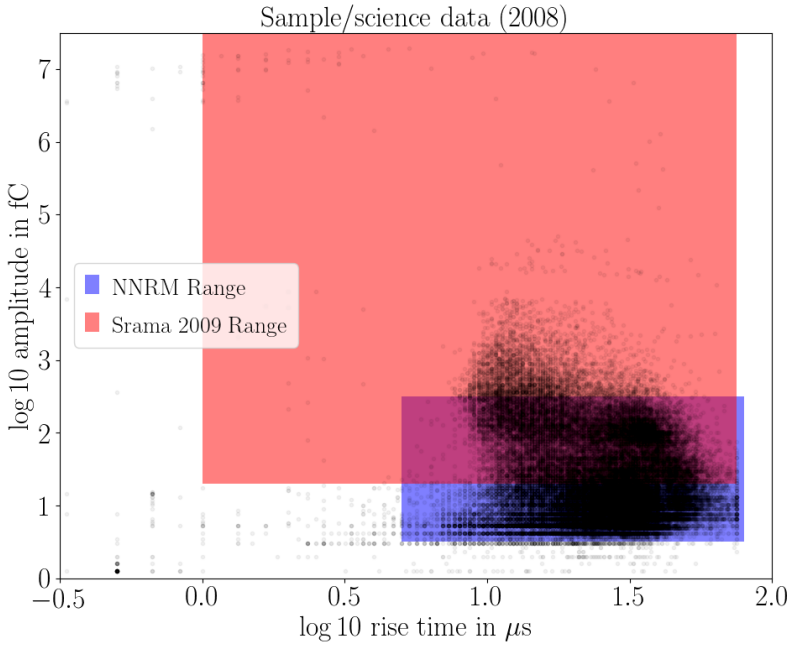


Figure 5.73.: Impact Ionisation Detector (IID) QI data from 2008 with the definition range of the Srama [2009] and the Neural Network Regression Model (NNRM) calibration functions. Each dot represents an actual measurement. The amplitude is plotted vs. the rise time. The red and blue rectangles indicate the definition range of the Srama [2009] and the Neural Network Regression Model (NNRM) functions; the purple rectangle is the overlapping domain.

than the new determined IID QI NNRM. However, this is due to the NNRM definition range that also covers very short rise times. A short sample time causing a rise time of only 1 or a few Digital Number (DN) might, however, not be suitable for further scientific evaluations. To compare the scientific subset, only data are considered that are in the definition range of both calibration functions (in total 22,000 data points).

To compare the effects of both calibration functions on the data set, the mean velocity and corresponding error determined by the NNRM is plotted vs. the Srama [2009] function (figure 5.74). The velocities are given in km/s and the color indicates the charge, respectively the amplitude of the detected particles. Same velocity values shall lay on the dashed lines that represent a simple linear fit.

A major part of the velocities determined by Srama [2009] is overestimated (w.r.t. to the new model). For velocities smaller than approximately 10 km/s the deviation of the determined velocities is less dominant for less charged particles than for higher charged ones. Afterwards, the deviations increases for the smaller charges and decreases for higher charged dust grains. The error of the particles is up to a factor of 4 improved with the NNRM.

The orbit database contains a significant large subset of Saturn- and Sun-unbound particles (eccentricity in ecliptic coordinates large or equal 1). Even for measurements within the E-Ring, where the Saturn-bound particles dominate the measurements, unbound particles are determined. The new models could un-bias these data properly since the new impact velocity values are by trend smaller.

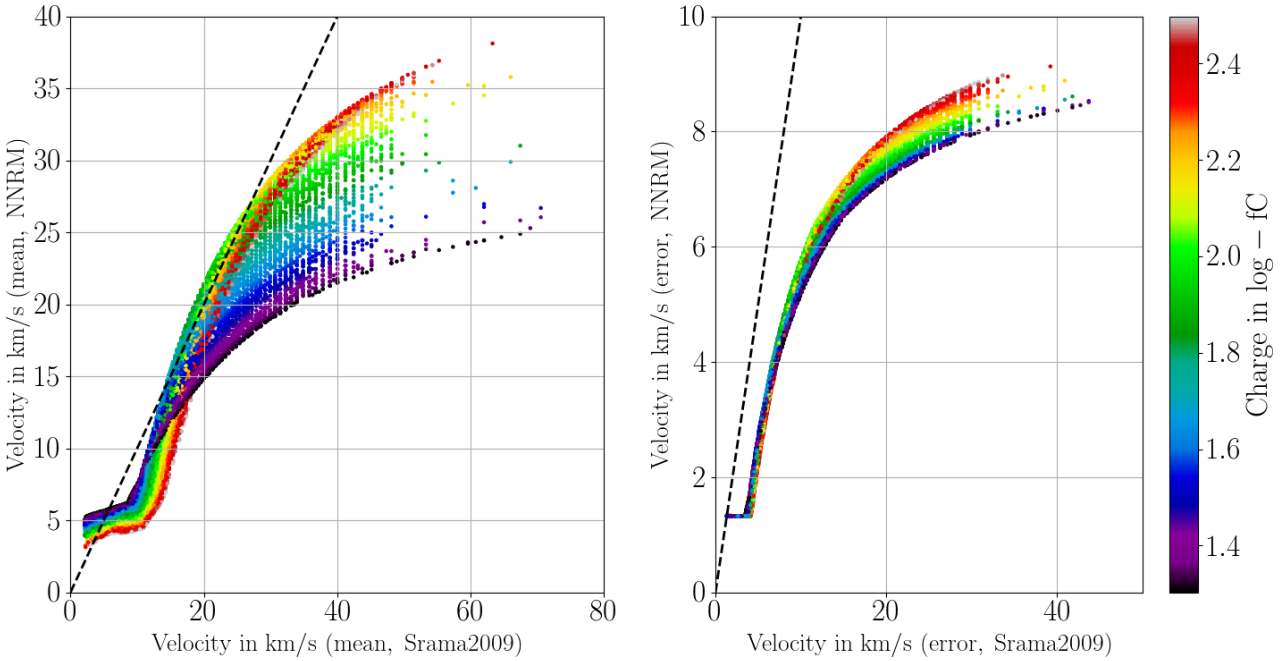


Figure 5.74.: Mean (left sub figure) and error (right figure) velocity determined by the Neural Network Regression Model (NNRM) vs. the results from the Srama [2009] calibration function in km/s. The dashed lines show a linear function indicating same results from both functions and are just a guide for the eye. The shows data are from the overlapping 2008 science data set and the color indicates the charge / amplitude of the measurement, respectively.

5.4.5. Conclusion and Outlook

This work has shown a new NN based approach to determine velocity calibration functions for the different CDA channels. Besides two functions for CAT and IID impacts a calibration function for wall impacts could be computed, too.

The new NN method appears to be sufficient and promising for further calibrations. The algorithms could be tested and used for the mass calibration, too, where hyperplane fits were applied by Srama [2000]. Also, the NN regression method can be used for future scientific instrument calibrations, e.g., for the Destiny Dust Analyser. The already developed models might be applicable for further 2-dimensional input data. A more sophisticated solution would be the direct implementation of NNs in the flight hardware to derive scientific parameters. Also, a chemical composition depending solution might be feasible, if according calibration data can be generated.

Considering the PEP8 coding standard, a PYTHON3 library was developed called CDACAL. The program stores all final network solutions and contains miscellaneous user-friendly functions to use the NNs as well as the already established functions by Srama et al. [2004]. The library is stored in the PYPI repository[10], and the contained documentation support the user to use the program properly.

[10] `https://pypi.python.org/pypi/cdacal`

5.5. Machine Learning based type-4 spectra classification

This section is about an ML based project to identify new, unknown type 4 spectra in the science database. Before this project 680 type 4 spectra were known, associated e.g., with dust from the ISD Altobelli et al. [2016]. A major part of recorded spectra in the Saturnian system have a water ice composition. The corresponding mass-spectra spectra are dominated by water molecule clusters (attached to a proton), resulting in a "comb-like" spectrum. Type 4 spectra have no water abundance and are associated with a mineral composition consisting of iron, kalium, calcium and others. All a-priori available spectra classifications have been performed by experts. This project shall support the Heidelberg working group to identify new type 4 data as well as demonstrating the feasibility of ML methods for classifications tasks. The project has been lead by the author of this thesis in cooperation with the students B. Knieling [Knieling, 2017] and F. Burkhardt [Burkhardt, 2017] who implemented the suggested algorithms for this classifications task: Support Vector Machine (SVM)s and K-Nearest Neighbor (KNN), respectively.

5.5.1. A-Priori classified data set

For the classification project 6936 mass spectra signals with corresponding labels were provided by F. Postberg's working group from the University of Heidelberg (priv. comm.). The data were classified manually in the last couple of years by miscellaneous students and employees. Furthermore, the labelled data cover dust impacts varying between 2004 and 2016. Different scientific projects required a classification of the signals, e.g., during ISD particle measurement campaigns, E-Ring analysis etc. Figure 5.75 shows the distribution of the provided data. Most spectra are water ice types (1 to 3). Around 10 % are type 4 mineral spectra and only 7 type 5 spectra are available. Fortunately 122 spectra are on hand to train an ML algorithm on noise data. If these data were not available, artificially generated noise signals could have been generated (e.g., with a white noise generator). Furthermore, different test pulses are considered from the science data base [Albin, 2014] that are correctly labelled.

This data set contains only uncompressed spectra with a length of 1018 data points (so called shrinking 1 signals). Further, around 2000 shrinking 2 spectra (half sized spectrum) are not considered. This project focuses on the compress-free signals (shrinking 1) without any possible compression-loss. A simple method to apply the shrinking 1 ML algorithm on shrinking 2 data is mentioned in the conclusion of this section.

To determine the spatial variety of the classified spectra, the complete dataset is separated in two sub-sets with shrinking 1 and 2. Figure 5.76 shows the spatial distribution of the shrinking 1 data. A top view on the Saturnian system shows the detection areas of the individual classes. The coordinate system in X and Y direction is given in SR. The inner light brown circle represents Saturn with its main ring (dark brown / grey ring). Enceladus' and Titan's orbit are shown as the inner and outer dashed circles. The different classes are color coded.

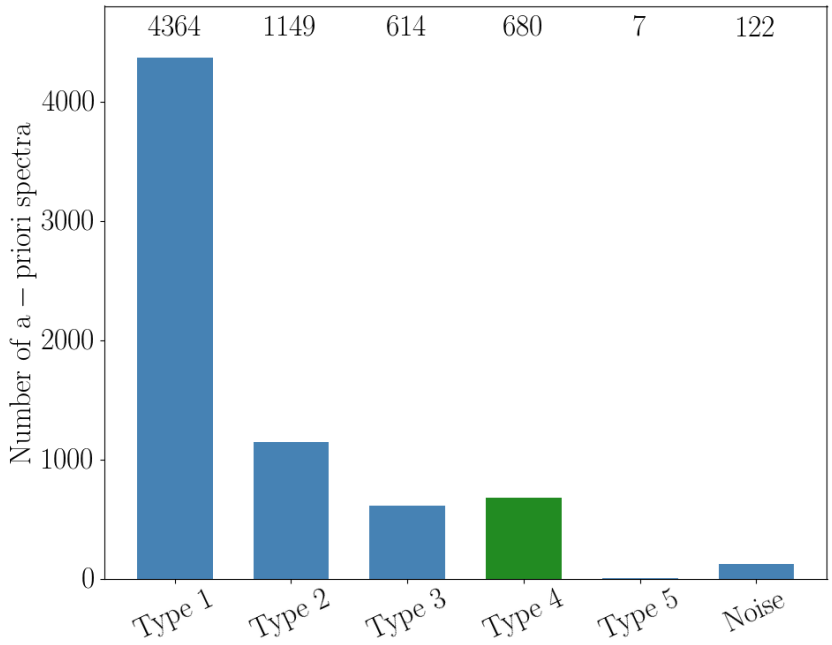


Figure 5.75.: A-priori classified spectra data set. 6 different QM signal types are available: Type 1 to 5 and a set of noise signals. The bars indicate the number of labelled signals (actual number above each bar). Since this project focuses on a 2 classification problem (type 4 - not type 4), the type 4 bar is differently coloured. In total, 6936 labelled data are available.

Most spectra classifications are within Titan's orbit, close to Enceladus and especially during periapsis phases. Thus, most spectra are classified within the E-Ring, resulting in type 1 and 2 and rarely type 3 spectra.

Another distinctive feature is the type 4 measurements beyond 20 SR. These detections are mainly from the ISD measurement campaign [Altobelli et al., 2016]. Type 4 spectra are also present within Titan's orbit.

Furthermore, the noise signals are present in all shown trajectory parts. However, especially one passage shows a high number of noise signals. Whether these are false classifications, anomal instrument behaviour or a large number of test pulses needs further investigation.

A part of the a-priori data have signal properties that are within the definition range of the velocity calibration functions and are therefore in the CDA DOCC database. However, only a fraction of these data are within a 95 % confidence interval of an orbit class. Around 3000 particles are Saturn bound particles, 500 are particles with $e \geq 1$ in HCEC and 20 are IDPs. 25 % are thus of exogenic origin and could origin from the mentioned velocity determination bias.

The idea of a possible velocity bias that later lead to the NNRM project to generate new velocity calibration functions, is constituted by considering the spatial detection distribution (figure 5.76) with an astrodynamical analysis.

Figure 5.77 shows the orbits of the Saturn bound particles. The types 1 to 4 are shown in the sub-figures a) to d) and visualize the median trajectories from a top-view perspective on the Saturnian system with Saturn and its ring in the center (in SCNRC). The inner and outer

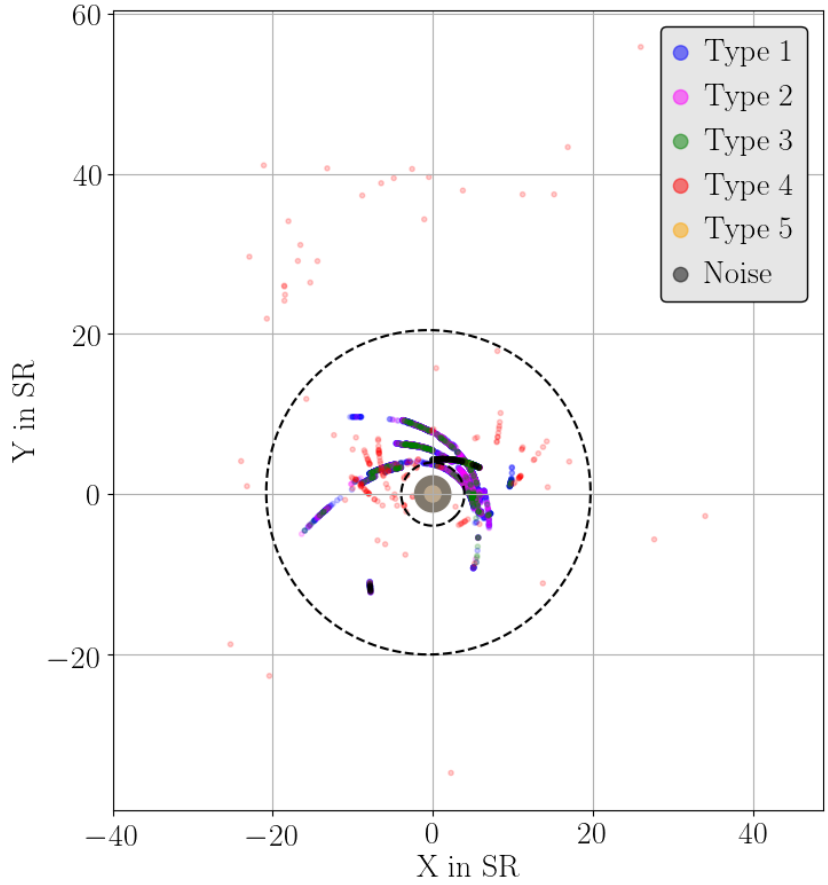


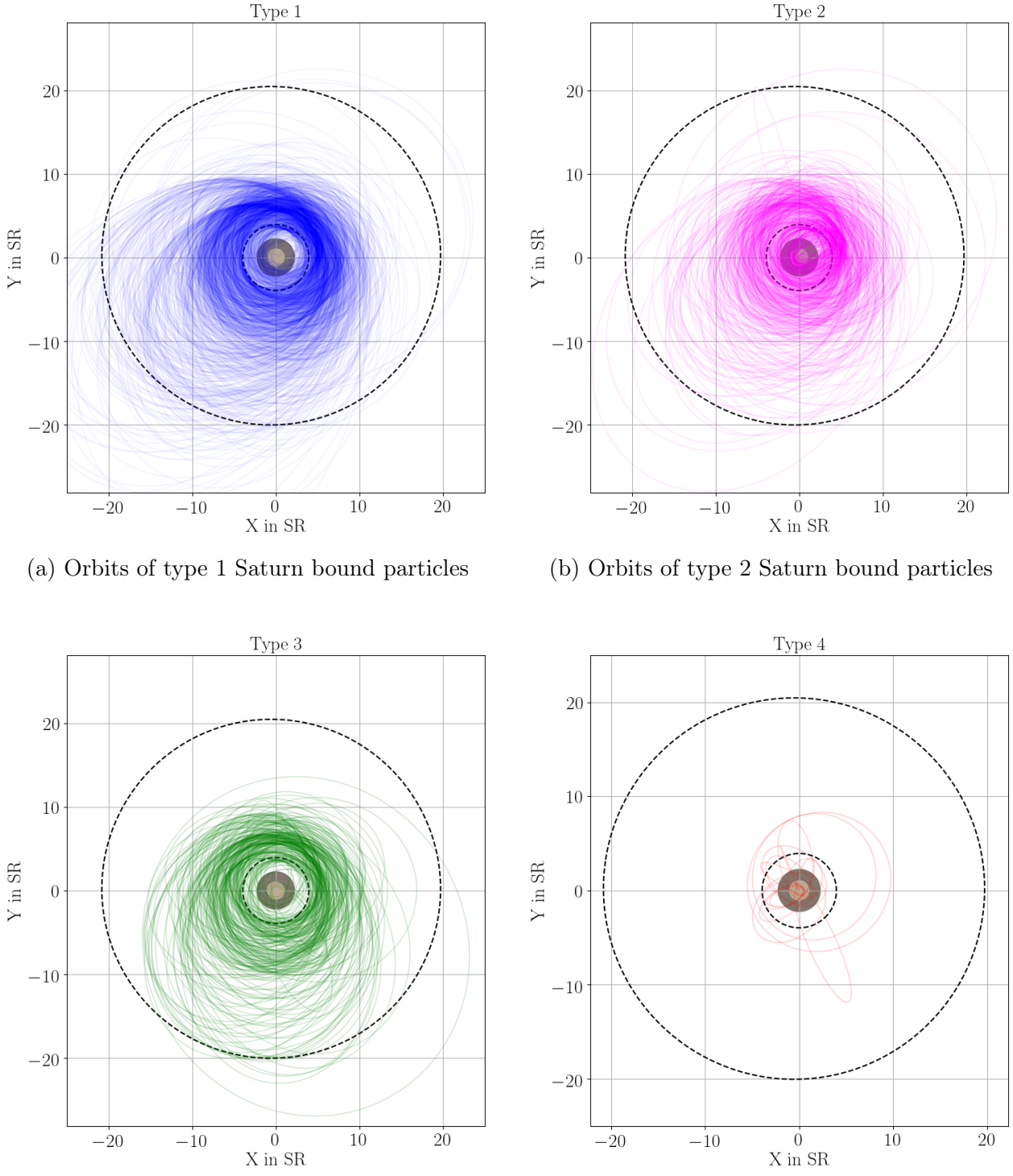
Figure 5.76.: Spatial detection distribution of the shrinking 1 spectra in the Saturnian system (top view). The X and Y coordinates are given in Saturn Radius / Radii (SR). In the center is Saturn (light brown) shown with its main ring (dark brown / grey). The inner and outer dashed circles represent the orbits of Enceladus and Titan, respectively. Each detection is shown as a dot in the plot, where each classification type is color coded and underlaid with a transparency.

dashed circles represent the orbits of Enceladus and Titan, respectively and the coordinates are given in SR. Each orbit is shown slightly transparent to enhance the densities of the orbits.

Most orbits of the type 1, 2 and type 3 particles revolve Saturn between 2 and 10 SR. For the fourth type only a few particles were detected. 2 show a circularly shaped orbit in the E-Ring, however, most type 4 particles are on very eccentric (and retrograde) orbits. The orbits of the water-ice grains show in total a stretched shape with high apsis parameters. The apsis medians are partly beyond the orbit of Titan.

Figure 5.76 shows the spatial distribution of Cassini's position for all detections. Comparing the orbits in the following figure with the spatial distribution it appears that for most particles the periapsis correlates with the detection position. This is visualised in figure 5.78, where the median periapsis is plotted vs. the distance of Cassini to Saturn in SR. The plot shows partly a linear correlation between both parameters indicating a velocity detection bias and / or a detection threshold bias to high energetic particles (high impact velocity). Particles on Keplerian orbits have a larger dwell time at the apsis than at the periapsis. Consequently, time-independently the detection probability of Keplerian-moving particles decreases for shorter

distances to Saturn. Thus, the correlation as shown in figure 5.78 is partly due to the mentioned velocity calibration bias.



(a) Orbits of type 1 Saturn bound particles

(b) Orbits of type 2 Saturn bound particles

(c) Orbits of type 3 Saturn bound particles

(d) Orbits of type 4 Saturn bound particles

Figure 5.77.: Orbits of the determined Saturn bound particles. Each subplot shows the orbits for each classified type. The coordinate system is Saturn-centric non-rotating coordinate system with Saturn and its ring in the center (topview). The inner and outer dashed circles represent the orbits of Enceladus and Titan, respectively. Edge-on views of the shown particle orbits is given in the appendix D.1.

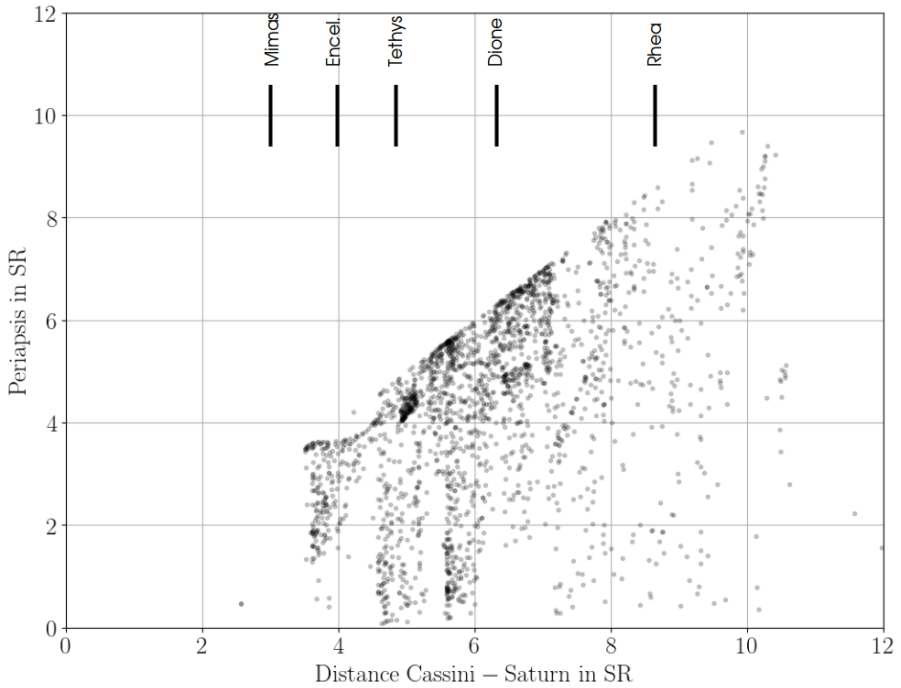


Figure 5.78.: Median periapsis of the detected shrinking 1 Saturn bound particles (before 2008) vs. the distance of the Cassini spacecraft to Saturn at the time of detection. The periapsis and distance are given in Saturn Radius / Radii (SR). 5 Saturnian moons are added, to visualize the distance coverage.

5.5.2. Training and verification

As already mentioned in chapter 2 a proper machine learning algorithm usage requires first a properly chosen ML algorithm, pre-processing and features extraction of the data and a proper 50/50 data splitting for training and verification purposes. Further, the required performance of the ML algorithms was defined by F. Postbergs working group (priv. comm.). The developed algorithms shall be considered as a "supportive tool" for the scientists. The classified spectra are manually analysed afterwards. The most important requirement is to minimise the number of false negatives (false negative rate). Unknown type 4 spectra should not be labelled as a non-type 4 ones. On the other hand, the false positive rate does not need to be strictly optimised to allow a more flexible development. Even with a total classification set of 50 % type 4 and non-type 4, the vast number of data would be reduced to a sub-set that can be investigated more easily.

Burkhardt [2017] and Knieling [2017] supported this project with the development of a feasible KNN and SVM algorithm, respectively. The results of both algorithms are shown in the following two sections.

5.5.3. KNN performance

The KNN algorithm was studied, tested and implemented by Burkhardt [2017]. Burkhardt focused especially on miscellaneous pre-processing and feature extraction methods and tools,

to determine the best performance for a KNN. Methods like Principle Component Analysis, Fast-Fourier Transformation, or dynamic time warping did not perform as well as more simple solutions, based on standardisation and normalisation. Tools from the library TSFRESH[11] that is optimised for big data signal processing, did not perform better. Quite the contrary, the computation time worse significantly.

The best performance was achieved with rather simple approaches. Burkhardt [2017] tested the KNN algorithm also on the raw data, normalised and / or standardised (miscellaneous functions) and used different metrics like the euclidean, chebyshev, or cityblock respectively manhattan metric. It was found that the normalised signals with only a few (less than 10) neighbours on the cityblock metric performed best. The true postive rate was 99.11 % on a half split verification set. The verification set contained 340 type 4 spectra and only 3, so around 1 % was wrongly classified by the algorithm. Only 13 of 3128 non-type 4 signals contaminated the data.

Thus a KNN based algorithm was developed for further classifications.

5.5.4. SVM performance

Knieling [2017] studied SVM based solutions and used filters to improve the classification results. With the standardisation and normalisation a slightly worse performance was achieved as with the KNN method. Further, Knieling applied Savitzky-Golay filters on the spectra signals to smooth the curves and reduce the noise of signals with less prominent features. He found that a window length of 41 and a polynom with 2 degrees performed well. The classification performance improved and the corresponding confusion matrix of the verification set was similar as determined by Burkhardt [2017]. The number of false negatives was 4, but the number of false positives was only 3. Further, the SVM algorithms used an RBF Kernel that needed also a hyper-parameter optimisation as shown in figure 5.79.

For further applications on the unlabelled data, two methods from Knieling's work were used: one SVM with only the standardised and normalised signals, and another one with the applied filter.

[11] `https://github.com/blue-yonder/tsfresh`, 12.04.2018

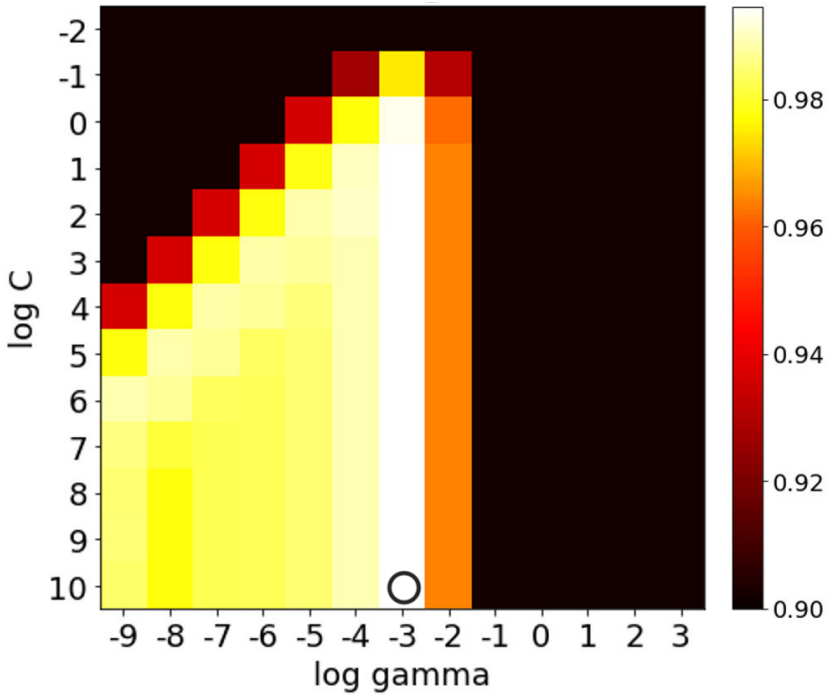


Figure 5.79.: Hyper-parameter optimisation for the Support Vector Machine (SVM) algorithm with an applied RBF-Kernel. The penalty parameter C (logarithmic) is plotted vs. the γ factor (logarithmic) and the corresponding color indicates the true positive rate. Modified: a circle indicates the best hyper-parameter set at $C = 10$ and $\gamma = 0.001$. From Knieling [2017].

5.5.5. Application on unknown data

3 developed models are then used and applied on the complete, unlabelled CDA database. The KNN algorithm with the cityblock metric and two SVM methods. One that only uses the standardised and normalised signals and a second one with applied Savitzky-Golay filter.

A classification data set is generated that contains the SCLK, and classification label of each method (type 4 or Non-type 4). Based on the three different classification results, a total prediction score is determined: 1/3 (one ML classified the signal as a type 4 spectrum), 2/3 (two out of three methods classified the signals as a type 4 spectrum) and 3/3 (all methods classified the same spectrum as type 4). The data have been send to C. Fischer from the Heidelberg working group, who classified all candidates by eye. In total, 2584 type 4 spectra have been found by the ML methods.

The visual inspection of the spectra reveal that 1021 spectra are actual type 4 spectra. Thus, a high error rate is present. Figure 5.80 shows a bar plot with the number of type 4 classifications by the ML methods and by F. Fischer (label: expert). The X axis shows how many methods (out of three) classify the corresponding signal as a type 4 spectrum. All methods together (3 of 3 algorithms label the same signal as a type 4 spectrum) have an overall performance of around 81 %.

To determine the individual performance of each algorithm the number of classifications

is considered vs. the corresponding number of type 4 identifications by the expert (figure 5.81). Relatively, the best performance is achieved by the scaled SVM algorithm. Most correct classifications are performed by the KNN method, the number of false positives however is significantly larger.

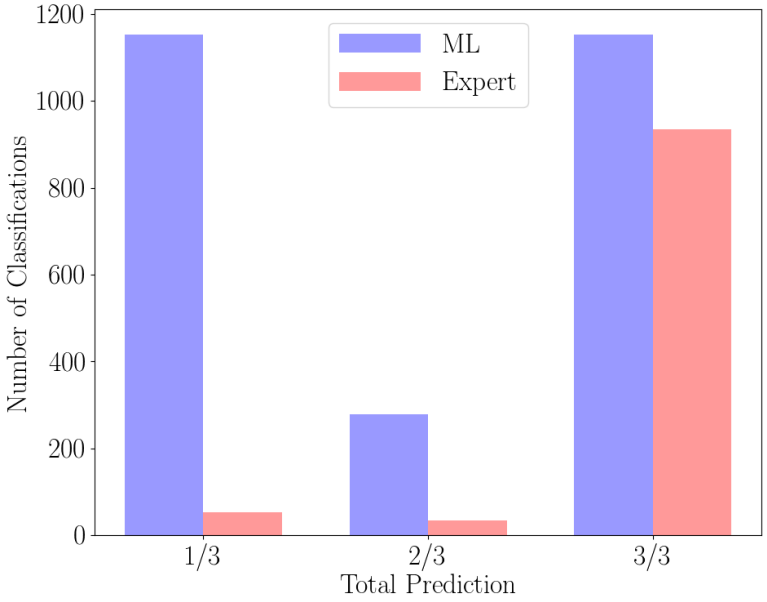


Figure 5.80.: Overall prediction performance. The bar plot shows the number of type 4 classifications based on 1, 2 and 3 (all) algorithms. Blue and red bars indicate the number of classification based on the Machine Learning (ML) algorithms and the expert, respectively. 81 % of all type 4 identifications that were found by all three methods correspond to actual type 4 data.

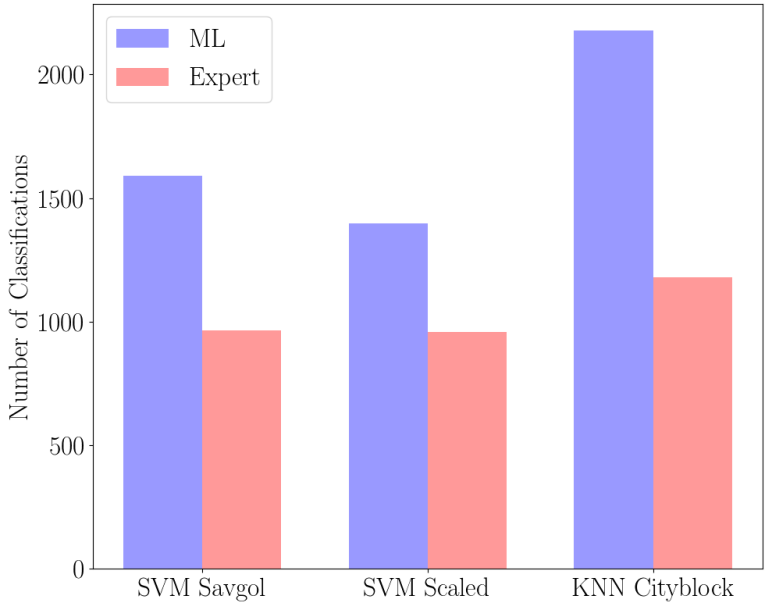


Figure 5.81.: Bar plot of the number of classifications for each individual classification method. Relatively, the scaled Support Vector Machine (SVM) method performed best. Most spectra were found however with the K-Nearest Neighbor (KNN) algorithm, though, the number of false positives is significantly higher.

5.5.6. Conclusion and Outlook

In this project ML based classifications tools were developed to support the Heidelberg working group in surveying the complete science database (hundred-thousands of unknown signals) for undiscovered type 4 spectra. With the developed models and in cooperation with C. Fischer 1021 new signals were found. The goal was to optimise the hyper-parameters of each method, to achieve an ideal true positive detection rate. Higher false positive rates were accepted as a drawback to allow a more flexible development. However, F. Postberg (priv. comm.) stated that finding all type 4 spectra had the highest priority and a contamination would be acceptable as long as the number of total spectra classifications would not complicate a following manually inspection.

The developed method of using all three algorithms in parallel lead to a satisfactory result. Based on the new unknown type 4 spectra, further dynamical and compositional analysis can be performed.

In future projects, NNs might improve the classification performance and allow a multi-class labelling of all data. Additionally, F. Willenborg (2019, in prep.) develops an unsupervised ML algorithm (Gaussian Mixture Model (GMM) and K-Means++) to distinguish and classify miscellaneous events considering all available channels (e.g. Noise, QI flares, IID or CAT impact, QA and QP grid impacts and possibly Wall impacts, too).

5.6. Conclusion and Outlook

In this chapter miscellaneous projects were presented that lead to new scientific insights or are a basis for further future research work. Similarly as the MC based orbit code for CILBO a multi central processing unit and PYTHON based program was developed to compute astrodynamical properties of CDA impact events. Considering the angular FOV and the velocity determination uncertainty, a MC program computes a detailed set of possible particle solutions, like e.g., Moon-originating coordinates, Tisserand parameters w.r.t. the largest moons, radiants information of ISD particles and IDPs as well as the orbital elements in different reference frames. Bötsch [2018] and Faber [2018] identified possible IDPs and candidates from ISD by applying the D_{SH} parameter from meteor science and graphs, as used in social media analysis.

An extensive test case of the DOCC database was the astrodynamical analysis of the ISD candidates found by Altobelli et al. [2016]. Here, a recent update on the upstream direction was provided, too; leading to the median upstream coordinates $\lambda = 263°$ and $\beta = 9°$ in EclipJ2000.

In cooperation with Knieling [2017] and Burkhardt [2017] over 1000 new type 4 spectra were found with machine learning algorithms. During this work, the dynamics of the already classified data were analysed. It was found that a possible velocity bias is present. This lead, in parallel, to the NNRM project, where 5 velocity functions were computed using Neural Networks. Also a feasible wall impact function was determined.

While working on this thesis, miscellaneous projects that are based on the concepts and ideas of this work are ongoing. First, a second DOCC database will be established with the NNRM velocity functions. Further, all developed tools are currently implemented in an overall PYTHON library that will be available over PIP. The orbit code, NNRMs, classification algorithms etc. are merged in cooperation with B. Knieling. To properly unbias observational directions a tool was developed that contains all pointing information of the Cassini mission. Considering the angular FOV of CDA and its targets, a KDE based tool generates a proper, normalised sky map to unfold observational biases. This will be used in the future to e.g. identify properly possible IDP streams or "meteor showers" into the Saturnian system (an internal report in this regard can be obtained from the author and would exhaust the scope of this thesis). Furthermore, a 1D convolutional deep learning network has been developed, tested and successfully applied on the mass spectra data. Based on all 30,000 manually classified spectra a KERAS based neural network classified all remaining spectra data in the database (classes: Noise, Type 1 to 5 and unknown / unidentified spectra) (Albin et al. 2019, in prep.).

Overall several tools and databases were developed with recent data analysis methods that allow further scientific research.

6. Conclusion and Outlook

6.1. Conclusion

This thesis was an extensive project driven work to determine if and how recent developments in data science can be applied in observational and in-situ cosmic dust research. During the work on this thesis two major data scientific fields were applied: Monte-Carlo (MC) based orbit computations and Machine Learning (ML) based data analysis. Two scientific projects were considered during this work: The Canary Island Long-Baseline Observatory (CILBO) and Cassini's on-board dust instrument, the Cosmic-Dust-Analyzer (CDA).

6.1.1. CILBO

With miscellaneous Kernel Density Estimator (KDE)s and other data structure describing methods different observational and technical biases were described in detail. Due to the wide field Field-of-view (FOV) of the stereoscopic camera system, a severe projection cut-off and vignetting cause a sensitivity drop to the edges of the Charged-Couple Device (CCD). Assuming an ideal lens a new metric was defined, the so called *detection efficiency*. A value of 1 indicates that the system detects all meteors that are within the limiting magnitude detection range of the system. However, the value lays for all systems between 0 and 1 and depends on the meteor's brightness as well as velocity. Faster meteors have a shorter pixel dwell time than slower ones. Individual pixels are exposed to less light and the Signal-To-Noise ratio (SNR) is worse. Fast meteors that appear brighter compensate on the other hand this effect. The analysis has shown that the *detection efficiency* is a multi-dimensional problem that effects e.g., the flux computation of meteor showers and sporadics.

Furthermore, miscellaneous effects were studied too, like e.g., the velocity and brightness depending accuracy determination. Bright and slow meteors overexpose pixels and the accuracy of the position determination is worse than for fast and fainter meteors. This effects directly the trajectory determination.

Considering time depending measurement biases it was found that several effects are due to a changing angular distance between camera boresight and the radiants of the sporadic Apex and Antihelion sources. Both cameras have an identical set-up: camera, lens, electronics, detection software etc. However Intensified CCD Camera (ICC)9 that is located on La Palma detects more fainter and slower meteors than its pendant on Teneriffe. The east-wards pointing La

Palma camera has a close angular distance to the Apex source in the second half of the night. Meteors from this region have thus a smaller apparent velocity and thus shorter pixel dwell time. The detection performance improves due to the better SNR.

A second project was the development of an orbit computation code, as part of an extension module for the Meteor Research Group's trajectory tool Meteor Orbit and Trajectory Software (MOTS). Based on stereoscopic measurements, velocity determination errors and astrometric uncertainties a MC tool was developed. With the current database over 10,000 orbits were computed and stored in an embedded database. First statistical analysis revealed e.g., the sporadic source regions. This project used bias conclusions from the previous project to properly handle e.g., velocity determination biases.

6.1.2. CDA

One of the major CDA projects was the development of the software tool Dust Orbit Computation Code (DOCC). The MC based program uses KDE based Probability Density Function (PDF)s for the velocity determination and the angular FOV of both targets to determine astro-dynamical properties of impact events within the Saturnian system. Using the SPICE toolkit an extensive database was computed that contains over a quarter million entries. Several tables consider different dynamical cases, like possible moon-bound / -out and -inbound solutions. Back propagating the particles in time, possible source coordinates can be determined of emitted dust particles e.g., from Enceladus. Orbital elements, and considering the Tisserand parameters w.r.t. the largest moons allow a detailed analysis of the E-Ring and discovery of possible ring structures. Interplanetary and interstellar dust particles are stored in separate tables, too. Using the median, Median Absolute Deviation (MAD), confidence interval etc. of the solutions lead to a statistical overview that will allow a detailed analysis.

A first application of the database was performed on the Interstellar Direction (ISD) particle candidates that were found by Altobelli et al. [2016]. One-third of the candidates are within the database and lead to a new determination of the interstellar upstream direction.

A noticeable amount of dust particles were classified as Sun-unbound particles. This leads to a new velocity calibration approach based on Neural Network (NN)s. For the two targets two new calibration functions were determined that are depending on the amplitude and rise time of the detections. The developed Neural Network Regression Model (NNRM) methods allowed the determination of a wall impact function, too.

Finally, supervised ML algorithms were developed to identify new, unknown type 4 spectra in the science database. Over 1000 spectra were successfully classified. This project proved that classification algorithms are a supportive and time saving way to handle large amount of data.

6.2. Outlook

This work was intentionally a compilation of different data scientific and Machine Learning based concepts that were applied on observational and in-situ dust detections. The algorithms were used for bias descriptions, unbiasing data, determining new calibration functions or to classify data sets. Kernel Density Estimator methods were used to derive Probability Density Functions for further Monte-Carlo based computations to determine astrodynamical properties of the Canary Island Long-Baseline Observatory (CILBO) and the Cosmic-Dust-Analyzer (CDA) measurements.

All concepts, the developed programs and the vast databases are a well basis for further extensive scientific and technical research.

6.2.1. Canary Island Long-Baseline Observatory (CILBO)

Miscellaneous technical and observational bias effect were already considered during the development of the orbit computation tool. However, especially the observational effects that were first identified and described in this work need further investigations. Scientific implications of these effects were mentioned, like e.g., a velocity and brightness depending meteor flux bias. Also, the mass index of meteor streams and sporadic sources might vary during an observation night due to the changing geometries between static camera boresight and radiants. These effects needs to be analysed in further numerical analysis to develop a proper (empirical) model of these effects. A comprehensive analysis could lead to an Machine Learning based un-biasing tool, where camera properties, detection time, boresight etc. are considered.

Also, the orbit database with over 10,000 orbits is a proper basis for further scientific and astro-dynamical analysis of meteors that were detected by the CILBO system. Currently, the orbit routines are embedded into the Meteor Orbit and Trajectory Software routines by the Meteor Research Group team and contractors. Further developments in the future will improve the computational performance, too.

Applying un-supervised Machine Learning algorithms on the orbit database could reveal further new scientific insights. This thesis has shown that Kernel Density Estimator methods are a sophisticated way to describe structures in data. Clustering algorithms could support future researchers in identifying yet undiscovered meteor showers.

6.2.2. Cosmic-Dust-Analyzer (CDA)

The developed Dust Orbit Computation Code database is an extensive basis for further scientific and also bias analysis. For example, Bötsch [2018] and Faber [2018] analysed dust particles of interplanetary and -stellar origin, respectively. They considered the D_{SH} parameter from meteor science and graphs (as used in the analysis of social media groups) to identify clusters / streams and the linkage to possible parent bodies. Furthermore, both used only

particles within a particle's rise time and amplitude range, where the velocity output from the Srama [2009] function overlaps with the results from the Neural Network Regression Model. The Neural Network Regression Model function will be implemented in the future to compute a second orbit database. Additionally, wall impacts could be considered, too.

Further scientific investigations could be: E-Ring dynamics, Enceladus plume dynamics, source mapping of emitted particles, search for particles linked to a possible Phoebe ring and also the identification of yet unknown ring structures. Also the analysis of the interplanetary and -stellar particles will need further research.

The Neural Network Regression Model method could be suitable for future calibration problems. With more data, also a spectrum or composition depending function could be derived. Analytical functions have the advantage of being explicitly defined, while Neural Networks need to be stored as objects on computers. However, more complex functions could benefit of this new approach. Possibly, with ML optimised hardware these tasks could be performed on future spacecrafts in real time.

The classification of CDA spectra was performed as a binary problem, labelling type 4 and non-type 4 spectra. With dimension reduction algorithms, un-supervised spectra clustering could improve the identification of "type 2.5" spectra and other yet unknown outliers. Currently, Willenborg (2019, in prep.) applies un-supervised clustering methods on all CDA channels (but the QM one), to identify actual impact events, noise signals, QI flares etc. A complete type classification of all mass spectra was performed by Albin et al. (2019, in prep.) with deep learning algorithms. They developed a KERAS based 1D convolutional neural network to classify the remaining, unknown mass spectra.

Furthermore, the author of this work develops in cooperation with B. Knieling a PYTHON library that will contain all described tools, algorithms and functions. CDACAL was a first attempt to provide a velocity calibration tool for the developed Neural Network Regression Models.

Bibliography

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

T. Albin. A global approach to analyse cassini cosmic dust analyser data using a relational data base system. Master's thesis, Max-Planck-Institut für Sonnensystemforschung, 2014.

T. Albin, D. Koschny, G. Drolshagen, R. Soja, B. Poppe, and R. Srama. De-biasing of the velocity determination for double station meteor observations from CILBO. In J.-L. Rault and P. Roggemans, editors, *Proceedings of the International Meteor Conference Mistelbach, Austria, 27-30 August 2015*, pages 214–219, January 2015a.

T. Albin, D. Koschny, G. Drolshagen, R. Soja, R. Srama, and B. Poppe. Influence of the pointing direction and detector sensitivity variations on the detection rate of a double station meteor camera. In J.-L. Rault and P. Roggemans, editors, *Proceedings of the International Meteor Conference Mistelbach, Austria, 27-30 August 2015*, pages 226–232, January 2015b.

T. Albin, D. Koschny, R. Soja, R. Srama, and B. Poppe. A Monte-Carlo based extension of the Meteor Orbit and Trajectory Software (MOTS) for computations of orbital elements. In A. Roggemans and P. Roggemans, editors, *International Meteor Conference Egmond, the Netherlands, 2-5 June 2016*, pages 20–25, January 2016.

T. Albin, D. Koschny, S. Molau, R. Srama, and B. Poppe. Analysis of the technical biases of meteor video cameras used in the cilbo system. *Geoscientific Instrumentation, Methods and Data Systems*, 6(1):125–140, 2017. doi: 10.5194/gi-6-125-2017.

N. Altobelli, S. Kempf, M. Landgraf, R. Srama, V. Dikarev, H. Krüger, G. Moragas-Klostermeyer, and E. Grün. Cassini between Venus and Earth: Detection of interstel-

lar dust. *Journal of Geophysical Research (Space Physics)*, 108:8032, October 2003. doi: 10.1029/2003JA009874.

N. Altobelli, S. Kempf, H. Krüger, M. Landgraf, M. Roy, and E. Grün. Interstellar dust flux measurements by the Galileo dust instrument between the orbits of Venus and Mars. *Journal of Geophysical Research (Space Physics)*, 110:A07102, July 2005. doi: 10.1029/2004JA010772.

N. Altobelli, F. Postberg, K. Fiege, M. Trieloff, H. Kimura, V. J. Sterken, H.-W. Hsu, J. Hillier, N. Khawaja, G. Moragas-Klostermeyer, J. Blum, M. Burton, R. Srama, S. Kempf, and E. Gruen. Flux and composition of interstellar dust at Saturn from Cassini's Cosmic Dust Analyzer. *Science*, 352:312–318, April 2016. doi: 10.1126/science.aac6397.

G. Andreev. The Influence of the Meteor Position on the Zenith Attraction. In D. Heinlein and D. Koschny, editors, *Proceedings of the International Meteor Conference, 9th IMC, Violau, Germany, 1990*, January 1991.

AndrewAnnex. `https://github.com/AndrewAnnex/SpiceyPy`, 2016.04.08.

Joseph Babcock. *Mastering Predictive Analytics with Python*. Packt Publishing, 2016. ISBN 978-1-78588-271-5.

W. J. Baggaley. Advanced Meteor Orbit Radar observations of interstellar meteoroids. J. Geophys. Res., 105:10353–10362, May 2000. doi: 10.1029/1999JA900383.

W. J. Baggaley, S. H. Marsh, and S. Close. Interstellar Meteors. *Dust in Planetary Systems*, 643:27–32, January 2007.

Roger R. Bate, Donald D. Mueller, and Jerry E. White. *Fundamentals of Astrodynamics (Dover Books on Aeronautical Engineering)*. Dover Publications, 1971. ISBN 0-486-60061-0.

U. Beckmann. *Dynamik von Staubteilchen in Saturns E-Ring*. PhD thesis, Ruprecht-Karls-Universitaet Heidelberg, 2008.

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2011. ISBN 0-387-31073-8.

Rodolfo Bonnin. *Building machine learning projects with TensorFlow : engaging projects that will teach you how complex data can be exploited to gain the most insight*. Packt Publishing, Birmingham, UK, 2016. ISBN 978-1-78646-658-7.

L. Bötsch. Astrodynamic analysis of cassini cda data: Interplanetary dust. Master's thesis, 2018.

Jo Bovy, David W. Hogg, and Sam T. Roweis. Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *Ann. Appl. Stat.*, 5(2B):1657–1677, 06 2011. doi: 10.1214/10-AOAS439. URL `https://doi.org/10.1214/10-AOAS439`.

N. Brosch, R. Helled, D. Polishook, E. Almoznino, and N. David. Meteor light curves: the relevant parameters. MNRAS, 355:111–119, November 2004. doi: 10.1111/j.1365-2966.2004.08300.x.

P. Brown, R. E. Spalding, D. O. ReVelle, E. Tagliaferri, and S. P. Worden. The flux of small near-Earth objects colliding with the Earth. Nature, 420:294–296, November 2002. doi: 10.1038/nature01238.

P. Brown, D. K. Wong, R. J. Weryk, and P. Wiegert. A meteoroid stream survey using the Canadian Meteor Orbit Radar. II: Identification of minor showers using a 3D wavelet transform. Icarus, 207:66–81, May 2010. doi: 10.1016/j.icarus.2009.11.015.

F. Burkhardt. A machine-learning based classification of cassini-cda spectra with k-nearest-neighbors, 2017.

John M. Chambers, William S. Cleveland, Beat Kleiner, and Paul A. Tukey. *Graphical Methods for Data Analysis (Statistics)*. Chapman and Hall/Cole Publishing Company, 1983. ISBN 978-0412052712.

François Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

J. E. Colwell, M. Horanyi, and E. Gren. Capture of Interplanetary and Interstellar Dust by the Jovian Magnetosphere. *Science*, 280:88, April 1998. doi: 10.1126/science.280.5360.88.

Alexander T. Combs. *Python Machine Learning Blueprints: Intuitive data projects you can relate to*. Packt Publishing, 2016. ISBN 978-1-78439-475-2.

Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000. ISBN 978-0-521-78019-3.

T. Denk, G. Neukum, T. Roatsch, C. C. Porco, J. A. Burns, G. G. Galuba, N. Schmedemann, P. Helfenstein, P. C. Thomas, R. J. Wagner, and R. A. West. Iapetus: Unique Surface Properties and a Global Color Dichotomy from Cassini Imaging. *Science*, 327:435, January 2010. doi: 10.1126/science.1177088.

V. Dikarev, E. Grün, J. Baggaley, D. Galligan, M. Landgraf, and R. Jehn. The new ESA meteoroid model. *Advances in Space Research*, 35:1282–1289, 2005. doi: 10.1016/j.asr.2005.05.014.

V. Dmitriev, V. Lupovka, and M. Gritsevich. A new approach to meteor orbit determination. In J.-L. Rault and P. Roggemans, editors, *Proceedings of the International Meteor Conference, Giron, France, 18-21 September 2014*, pages 157–159, February 2014.

M. K. Dougherty, L. W. Esposito, and S. M. Krimigis. *Saturn from Cassini-Huygens*. Springer, 2009. ISBN 978-1-4020-9217-6.

G. Drolshagen, J.A.M. McDonnell, T. Stevenson, R. Aceti, and L. Gerlach. Post-flight measurements of meteoroid/debris impact features on eureca and the hubble solar array. *Advances in Space Research*, 16(11):85 – 89, 1995. ISSN 0273-1177. doi: https://doi.org/10.1016/0273-1177(95)98756-E. Space Debris.

Gerhard Drolshagen, Detlef Koschny, Sandra Drolshagen, Jana Kretschmer, and BjÃűrn Poppe. Mass accumulation of earth from interplanetary dust, meteoroids, asteroids and comets. *Planetary and Space Science*, 143:21 – 27, 2017. ISSN 0032-0633. doi: https://doi.org/10.1016/j.pss.2016.12.010. SI:Meteoroids 2016.

J. D. Drummond. A test of comet and meteor shower associations. Icarus, 45:545–553, March 1981. doi: 10.1016/0019-1035(81)90020-8.

M. Faber. Astrodynamische analyse von cassini cda messungen: Interstellarer staub. Master's thesis, 2018.

D. P. Galligan. Performance of the D-criteria in recovery of meteoroid stream orbits in a radar data set. MNRAS, 327:623–628, October 2001. doi: 10.1046/j.1365-8711.2001.04858.x.

J. R. Goeller and E. Grün. Calibration of the Galileo/Ulysses dust detectors with different projectile materials and at varying impact angles. Planet. Space Sci., 37:1197–1206, October 1989. doi: 10.1016/0032-0633(89)90014-7.

G.A. Graham, N. McBride, A.T. Kearsley, G. Drolshagen, S.F. Green, J.A.M. McDonnell, M.M. Grady, and I.P. Wright. The chemistry of micrometeoroid and space debris remnants captured on hubble space telescope solar cells. *International Journal of Impact Engineering*, 26(1):263 – 274, 2001. ISSN 0734-743X. doi: https://doi.org/10.1016/S0734-743X(01)00087-2.

E. Grün, H. A. Zook, H. Fechtig, and R. H. Giese. Collisional balance of the meteoritic complex. Icarus, 62:244–272, May 1985. doi: 10.1016/0019-1035(85)90121-6.

E. Grün, H. Fechtig, M. S. Hanner, J. Kissel, B.-A. Lindblad, D. Linkert, D. Maas, G. E. Morfill, and H. A. Zook. The Galileo Dust Detector. Space Sci. Rev., 60:317–340, May 1992. doi: 10.1007/BF00216860.

E. Grün, H. A. Zook, M. Baguhl, A. Balogh, S. J. Bame, H. Fechtig, R. Forsyth, M. S. Hanner, M. Horanyi, J. Kissel, B.-A. Lindblad, D. Linkert, G. Linkert, I. Mann, J. A. M. McDonnell,

G. E. Morfill, J. L. Phillips, C. Polanskey, G. Schwehm, N. Siddique, P. Staubach, J. Svestka, and A. Taylor. Discovery of Jovian dust streams and interstellar grains by the ULYSSES spacecraft. Nature, 362:428–430, April 1993. doi: 10.1038/362428a0.

E. Grün, M. Baguhl, D. P. Hamilton, J. Kissel, D. Linkert, G. Linkert, and R. Riemann. Reduction of Galileo and Ulysses dust data. Planet. Space Sci., 43:941–951, August 1995. doi: 10.1016/0032-0633(94)00232-G.

P. S. Gural. A new method of meteor trajectory determination applied to multiple unsynchronized video cameras. *Meteoritics and Planetary Science*, August 2012. doi: 10.1111/j.1945-5100.2012.01402.x.

C. J. Hansen, L. Esposito, A. I. F. Stewart, J. Colwell, A. Hendrix, W. Pryor, D. Shemansky, and R. West. Enceladus' Water Vapor Plume. *Science*, 311:1422–1425, March 2006. doi: 10.1126/science.1121254.

J. K. Hillier, S. F. Green, N. McBride, N. Altobelli, F. Postberg, S. Kempf, J. Schwanethal, R. Srama, J. A. M. McDonnell, and E. Grün. Interplanetary dust detected by the Cassini CDA Chemical Analyser. Icarus, 190:643–654, October 2007. doi: 10.1016/j.icarus.2007.03.024.

M. Horanyi, J. A. Burns, and D. P. Hamilton. The dynamics of Saturn's E ring particles. Icarus, 97:248–259, June 1992. doi: 10.1016/0019-1035(92)90131-P.

H.-W. Hsu, S. Kempf, and C. M. Jackman. Observation of saturnian stream particles in the interplanetary space. Icarus, 206:653–661, April 2010. doi: 10.1016/j.icarus.2009.06.033.

H.-W. Hsu, S. Kempf, F. Postberg, M. Trieloff, M. Burton, M. Roy, G. Moragas-Klostermeyer, and R. Srama. Cassini dust stream particle measurements during the first three orbits at Saturn. *Journal of Geophysical Research (Space Physics)*, 116:A08213, August 2011. doi: 10.1029/2010JA015959.

H.-W. Hsu, F. Postberg, Y. Sekine, T. Shibuya, S. Kempf, M. Horányi, A. Juhász, N. Altobelli, K. Suzuki, Y. Masaki, T. Kuwatani, S. Tachibana, S.-I. Sirono, G. Moragas-Klostermeyer, and R. Srama. Ongoing hydrothermal activities within Enceladus. Nature, 519:207–210, March 2015. doi: 10.1038/nature14262.

Guang-Bin Huang. Learning capability and storage capacity of two-hidden-layer feedforward networks. *Trans. Neur. Netw.*, 14(2):274–281, March 2003. ISSN 1045-9227. doi: 10.1109/TNN.2003.809401. URL `http://dx.doi.org/10.1109/TNN.2003.809401`.

Mia Hubert and Michiel Debruyne. Minimum covariance determinant. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):36–43, dec 2009. doi: 10.1002/wics.61.

Ž. Ivezić, A.J. Connolly, J.T. Vanderplas, and A. Gray. *Statistics, Data Mining and Machine Learning in Astronomy*. Princeton University Press, 2014.

Peter Jenniskens. *Meteor Showers and their Parent Comets*. Cambridge University Press, 2008. ISBN 978-0-521-07635-7.

D. Jewitt, H. Hsieh, and J. Agarwal. *The Active Asteroids*, pages 221–241. 2015. doi: 10.2458/azu_uapress_9780816532131-ch012.

G. H. Jones, E. Roussos, N. Krupp, U. Beckmann, A. J. Coates, F. Crary, I. Dandouras, V. Dikarev, M. K. Dougherty, P. Garnier, C. J. Hansen, A. R. Hendrix, G. B. Hospodarsky, R. E. Johnson, S. Kempf, K. K. Khurana, S. M. Krimigis, H. Krüger, W. S. Kurth, A. Lagg, H. J. McAndrews, D. G. Mitchell, C. Paranicas, F. Postberg, C. T. Russell, J. Saur, M. Seiß, F. Spahn, R. Srama, D. F. Strobel, R. Tokar, J.-E. Wahlund, R. J. Wilson, J. Woch, and D. Young. The Dust Halo of Saturn's Largest Icy Moon, Rhea. *Science*, 319:1380, March 2008. doi: 10.1126/science.1151524.

T. J. Jopek. Remarks on the meteor orbital similarity D-criterion. Icarus, 106:603, December 1993. doi: 10.1006/icar.1993.1195.

S. Kempf, R. Srama, N. Altobelli, S. Auer, V. Tschernjawski, J. Bradley, M. E. Burton, S. Helfert, T. V. Johnson, H. Krüger, G. Moragas-Klostermeyer, and E. Grün. Cassini between Earth and asteroid belt: first in-situ charge measurements of interplanetary grains. Icarus, 171:317–335, October 2004. doi: 10.1016/j.icarus.2004.05.017.

S. Kempf, R. Srama, M. Horányi, M. Burton, S. Helfert, G. Moragas-Klostermeyer, M. Roy, and E. Grün. High-velocity streams of dust originating from Saturn. Nature, 433:289–291, January 2005. doi: 10.1038/nature03218.

S. Kempf, U. Beckmann, G. Moragas-Klostermeyer, F. Postberg, R. Srama, T. Economou, J. Schmidt, F. Spahn, and E. Grün. The E ring in the vicinity of Enceladus. I. Spatial distribution and properties of the ring particles. Icarus, 193:420–437, February 2008. doi: 10.1016/j.icarus.2007.06.027.

B. Knieling. A machine-learning based classification of cassini-cda spectra with support-vector-machines, 2017.

D. Koschny and J. Diaz del Rio. Meteor Orbit and Trajectory Software (MOTS) - Determining the Position of a Meteor with Respect to the Earth Using Data Collected with the Software MetRec. *WGN, Journal of the International Meteor Organization*, 30:87–101, August 2002.

D. Koschny, F. Bettonvil, J. Licandro, C. v. d. Luijt, J. Mc Auliffe, H. Smit, H. Svedhem, F. de Wit, O. Witasse, and J. Zender. A double-station meteor camera set-up in the canary

islands - cilbo. *Geoscientific Instrumentation, Methods and Data Systems*, 2(2):339–348, 2013. doi: 10.5194/gi-2-339-2013.

P. Koten and J. Borovička. Light curves of faint meteors. In B. Warmbein, editor, *Meteoroids 2001 Conference*, volume 495 of *ESA Special Publication*, pages 259–264, November 2001.

P. Koten, J. Borovička, P. Spurný, H. Betlem, and S. Evans. Atmospheric trajectories and light curves of shower meteors. A&A, 428:683–690, December 2004. doi: 10.1051/0004-6361:20041485.

Ľ. Kresák. Relation of meteor orbits to the orbits of comets and asteroids. *Smithsonian Contributions to Astrophysics*, 11:9, 1967.

A. V. Krivov, H. Krüger, E. Grün, K.-U. Thiessenhusen, and D. P. Hamilton. A tenuous dust ring of Jupiter formed by escaping ejecta from the Galilean satellites. *Journal of Geophysical Research (Planets)*, 107:5002, January 2002. doi: 10.1029/2000JE001434.

H. Krüger, E. Grün, A. Graps, D. Bindschadler, S. Dermott, H. Fechtig, B. A. Gustafson, D. P. Hamilton, M. S. Hanner, M. Horányi, J. Kissel, B. A. Lindblad, D. Linkert, G. Linkert, I. Mann, J. A. M. McDonnell, G. E. Morfill, C. Polanskey, G. Schwehm, R. Srama, and H. A. Zook. One year of Galileo dust data from the Jovian system: 1996. Planet. Space Sci., 49: 1285–1301, November 2001. doi: 10.1016/S0032-0633(01)00053-8.

H. Krüger, D. Bindschadler, S. F. Dermott, A. L. Graps, E. Grün, B. A. Gustafson, D. P. Hamilton, M. S. Hanner, M. Horányi, J. Kissel, B. A. Lindblad, D. Linkert, G. Linkert, I. Mann, J. A. M. McDonnell, R. Moissl, G. E. Morfill, C. Polanskey, G. Schwehm, R. Srama, and H. A. Zook. Galileo dust data from the jovian system: 1997 1999. Planet. Space Sci., 54:879–910, August 2006. doi: 10.1016/j.pss.2006.04.010.

H. Krüger, D. P. Hamilton, R. Moissl, and E. Grün. Galileo in-situ dust measurements in Jupiter's gossamer rings. Icarus, 203:198–213, September 2009. doi: 10.1016/j.icarus.2009.03.040.

H. Krüger, D. Bindschadler, S. F. Dermott, A. L. Graps, E. Grün, B. A. Gustafson, D. P. Hamilton, M. S. Hanner, M. Horányi, J. Kissel, D. Linkert, G. Linkert, I. Mann, J. A. M. McDonnell, R. Moissl, G. E. Morfill, C. Polanskey, M. Roy, G. Schwehm, and R. Srama. Galileo dust data from the jovian system: 2000 to 2003. Planet. Space Sci., 58:965–993, June 2010. doi: 10.1016/j.pss.2010.03.003.

H. Krüger, P. Strub, E. Grün, and V. J. Sterken. Sixteen Years of Ulysses Interstellar Dust Measurements in the Solar System. I. Mass Distribution and Gas-to-dust Mass Ratio. ApJ, 812:139, October 2015. doi: 10.1088/0004-637X/812/2/139.

M. Landgraf. Modeling the motion and distribution of interstellar dust inside the heliosphere. J. Geophys. Res., 105:10303–10316, May 2000. doi: 10.1029/1999JA900243.

K. G. Malmquist. On some relations in stellar statistics. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100:1–52, March 1922.

R. E. McCrosky. The meteor wake. AJ, 63:97, March 1958. doi: 10.1086/107698.

Ernst Messerschmid and Stefanos Fasoulas. *Raumfahrtsysteme: Eine Einführung mit Übungen und Lösungen (German Edition)*. Springer Vieweg, 2017. ISBN 978-3-662-49637-4.

S. Molau. The Meteor Detection Software METREC. In R. Arlt and A. Knoefel, editors, *Proceedings of the International Meteor Conference, 17th IMC, Stara Lesna, Slovakia, 1998*, pages 9–16, January 1999.

S. Molau and M. Nitschke. Computer-Based Meteor Search: a New Dimension in Video Meteor Observation. *WGN, Journal of the International Meteor Organization*, 24:119–123, August 1996.

Carl D. Murray and Stanley F. Dermott. *Solar System Dynamics*. Cambridge University Press, 2000. ISBN 978-0-521-57597-3.

I. S. Murray, R. L. Hawkes, and P. Jenniskens. Airborne intensified charge-coupled device observations of the 1998 Leonid shower. *Meteoritics and Planetary Science*, 34:949–958, November 1999. doi: 10.1111/j.1945-5100.1999.tb01413.x.

NASA/JPL, a. `http://saturn.jpl.nasa.gov/spacecraft/cassiniorbiterinstruments/instrumentscassinicda/`, 2016.02.09.

NASA/JPL, b. `http://ssd.jpl.nasa.gov/?sat_elem`, 2016.03.31.

NASA/NSSDC. `http://nssdc.gsfc.nasa.gov/planetary/planetfact.html`, 2016.03.31.

Francis Nimmo, Carolyn Porco, and Colin Mitchell. Tidally modulated eruptions on enceladus: Cassini iss observations and models. *The Astronomical Journal*, 148(3):46, 2014.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Christoph Polle. Entwicklung eines numerischen integrators zur berechnung von orbits innerhalb des erde-mond-systems, 2017.

C. C. Porco, P. Helfenstein, P. C. Thomas, A. P. Ingersoll, J. Wisdom, R. West, G. Neukum, T. Denk, R. Wagner, T. Roatsch, S. Kieffer, E. Turtle, A. McEwen, T. V. Johnson, J. Rathbun, J. Veverka, D. Wilson, J. Perry, J. Spitale, A. Brahic, J. A. Burns, A. D. Del Genio, L. Dones, C. D. Murray, and S. Squyres. Cassini Observes the Active South Pole of Enceladus. *Science*, 311:1393–1401, March 2006. doi: 10.1126/science.1123013.

Carolyn Porco, Daiana DiNino, and Francis Nimmo. How the geysers, tidal stresses, and thermal emission across the south polar terrain of enceladus are related. *The Astronomical Journal*, 148(3):45, 2014.

F. Postberg, S. Kempf, R. Srama, S. F. Green, J. K. Hillier, N. McBride, and E. Grün. Composition of jovian dust stream particles. Icarus, 183:122–134, July 2006. doi: 10.1016/j.icarus.2006.02.001.

F. Postberg, S. Kempf, J. K. Hillier, R. Srama, S. F. Green, N. McBride, and E. Grün. The E-ring in the vicinity of Enceladus. II. Probing the moon's interior—The composition of E-ring particles. Icarus, 193:438–454, February 2008. doi: 10.1016/j.icarus.2007.09.001.

F. Postberg, S. Kempf, J. Schmidt, N. Brilliantov, A. Beinsen, B. Abel, U. Buck, and R. Srama. Sodium salts in E-ring ice grains from an ocean below the surface of Enceladus. Nature, 459: 1098–1101, June 2009. doi: 10.1038/nature08046.

F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, pages 65–386, 1958.

Peter J. Rousseeuw and Katrien Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, aug 1999. doi: 10.1080/00401706.1999.10485670.

F. Scipioni, F. Tosi, K. Stephan, G. Filacchione, M. Ciarniello, F. Capaccioni, and P. Cerroni. Spectroscopic classification of icy satellites of Saturn I: Identification of terrain units on Dione. Icarus, 226:1331–1349, November 2013. doi: 10.1016/j.icarus.2013.08.008.

F. Scipioni, F. Tosi, K. Stephan, G. Filacchione, M. Ciarniello, F. Capaccioni, and P. Cerroni. Spectroscopic classification of icy satellites of Saturn II: Identification of terrain units on Rhea. Icarus, 234:1–16, May 2014. doi: 10.1016/j.icarus.2014.02.010.

D.W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley Series in Probability and Statistics. Wiley, 2015. ISBN 9781118575536.

H. T. Smith, M. Shappirio, R. E. Johnson, D. Reisenfeld, E. C. Sittler, F. J. Crary, D. J. McComas, and D. T. Young. Enceladus: A potential source of ammonia products and molecular nitrogen for Saturn's magnetosphere. *Journal of Geophysical Research (Space Physics)*, 113:A11206, November 2008. doi: 10.1029/2008JA013352.

R. H. Soja, D. P. Hamilton, and N. Altobelli. A new analysis of Galileo dust data near Jupiter. Planet. Space Sci., 109:76–91, May 2015a. doi: 10.1016/j.pss.2015.01.020.

R. H. Soja, J. T. Herzog, M. Sommer, J. Rodmann, J. Vaubaillon, P. Strub, T. Albin, V. Sterken, A. Hornig, L. Bausch, E. Grün, and R. Srama. Meteor storms and showers with the IMEX model. In J.-L. Rault and P. Roggemans, editors, *International Meteor Conference Mistelbach, Austria*, pages 66–69, January 2015b.

R. H. Soja, M. Sommer, J. Herzog, J. Agarwal, J. Rodmann, R. Srama, J. Vaubaillon, P. Strub, A. Hornig, L. Bausch, and E. Grün. Characteristics of the dust trail of 67P/Churyumov-Gerasimenko: an application of the IMEX model. A&A, 583:A18, November 2015c. doi: 10.1051/0004-6361/201526184.

R. B. Southworth and G. S. Hawkins. Statistics of meteor streams. *Smithsonian Contributions to Astrophysics*, 7:261, 1963.

F. Spahn, J. Schmidt, N. Albers, M. Hörning, M. Makuch, M. Seiß, S. Kempf, R. Srama, V. Dikarev, S. Helfert, G. Moragas-Klostermeyer, A. V. Krivov, M. Sremčević, A. J. Tuzzolino, T. Economou, and E. Grün. Cassini Dust Measurements at Enceladus and Implications for the Origin of the E Ring. *Science*, 311:1416–1418, March 2006. doi: 10.1126/science.1121375.

J. R. Spencer and T. Denk. Formation of Iapetus' Extreme Albedo Dichotomy by Exogenically Triggered Thermal Ice Migration. *Science*, 327:432, January 2010. doi: 10.1126/science.1177132.

J. R. Spencer, J. C. Pearl, M. Segura, F. M. Flasar, A. Mamoutkine, P. Romani, B. J. Buratti, A. R. Hendrix, L. J. Spilker, and R. M. C. Lopes. Cassini Encounters Enceladus: Background and the Discovery of a South Polar Hot Spot. *Science*, 311:1401–1405, March 2006. doi: 10.1126/science.1121661.

J. N. Spitale, T. A. Hurford, A. R. Rhoden, E. E. Berkson, and S. S. Platts. Curtain eruptions from Enceladus' south-polar terrain. Nature, 521:57–60, May 2015. doi: 10.1038/nature14368.

R. Srama. *Vom Cosmic-Dust-Analyzer zur Modellbeschreibung wissenschaftlicher Raumsonden.* PhD thesis, Fachgebiet Raumfahrttechnik Technische Universität München, 2000.

R. Srama. *Cassini-Huygens and Beyond – Tools for Dust Astronomy.* Habilitation, Fakultät für Luft- und Raumfahrttechnik und Geodäsie der Universität Stuttgart, 2009.

R. Srama. In situ dust measurements by the Cassini Cosmic Dust Analyzer in 2014 and beyond. *European Planetary Science Congress 2015, held 27 September - 2 October, 2015 in Nantes, France, id.EPSC2015-846*, 10:EPSC2015-846, October 2015.

R. Srama, T. J. Ahrens, N. Altobelli, S. Auer, J. G. Bradley, M. Burton, V. V. Dikarev, T. Economou, H. Fechtig, M. Görlich, M. Grande, A. Graps, E. Grün, O. Havnes, S. Helfert, M. Horanyi, E. Igenbergs, E. K. Jessberger, T. V. Johnson, S. Kempf, A. V. Krivov, H. Krüger, A. Mocker-Ahlreep, G. Moragas-Klostermeyer, P. Lamy, M. Landgraf, D. Linkert, G. Linkert, F. Lura, J. A. M. McDonnell, D. Möhlmann, G. E. Morfill, M. Müller, M. Roy, G. Schäfer, G. Schlotzhauer, G. H. Schwehm, F. Spahn, M. Stübig, J. Svestka, V. Tschernjawski, A. J. Tuzzolino, R. Wäsch, and H. A. Zook. The Cassini Cosmic Dust Analyzer. Space Sci. Rev., 114:465–518, September 2004. doi: 10.1007/s11214-004-1435-z.

R. Srama, S. Kempf, G. Moragas-Klostermeyer, S. Helfert, T. J. Ahrens, N. Altobelli, S. Auer, U. Beckmann, J. G. Bradley, M. Burton, V. V. Dikarev, T. Economou, H. Fechtig, S. F. Green, M. Grande, O. Havnes, J. K. Hillier, M. Horanyi, E. Igenbergs, E. K. Jessberger, T. V. Johnson, H. Krüger, G. Matt, N. McBride, A. Mocker, P. Lamy, D. Linkert, G. Linkert, F. Lura, J. A. M. McDonnell, D. Möhlmann, G. E. Morfill, F. Postberg, M. Roy, G. H. Schwehm, F. Spahn, J. Svestka, V. Tschernjawski, A. J. Tuzzolino, R. Wäsch, and E. Grün. In situ dust measurements in the inner Saturnian system. Planet. Space Sci., 54:967–987, August 2006. doi: 10.1016/j.pss.2006.05.021.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. URL `http://jmlr.org/papers/v15/srivastava14a.html`.

V. J. Sterken, P. Strub, H. Krüger, R. von Steiger, and P. Frisch. Sixteen Years of Ulysses Interstellar Dust Measurements in the Solar System. III. Simulations and Data Unveil New Insights into Local Interstellar Dustkrueger2015. ApJ, 812:141, October 2015. doi: 10.1088/0004-637X/812/2/141.

Edward Stoken. Fragmentation and wake formation in faint meteors: Implications for the structure and ablation of small meteoroids, 2014.

P. Strub, H. Krüger, and V. J. Sterken. Sixteen Years of Ulysses Interstellar Dust Measurements in the Solar System. II. Fluctuations in the Dust Flow from the Data. ApJ, 812:140, October 2015. doi: 10.1088/0004-637X/812/2/140.

D. Subasinghe, M. D. Campbell-Brown, and E. Stokan. Physical characteristics of faint meteors by light curve and high-resolution observations, and the implications for parent bodies. MNRAS, 457:1289–1298, April 2016. doi: 10.1093/mnras/stw019.

A. D. Taylor and W. G. Elford. Meteoroid orbital element distributions at 1 AU deduced from the Harvard Radio Meteor Project observations. *Earth, Planets, and Space*, 50:569–575, June 1998. doi: 10.1186/BF03352150.

Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001. doi: 10.1111/1467-9868.00293.

M. S. Tiscareno, J. A. Burns, J. N. Cuzzi, and M. M. Hedman. Cassini imaging search rules out rings around Rhea. Geophys. Res. Lett., 37:L14205, July 2010. doi: 10.1029/2010GL043663.

J.W. Tukey. *Exploratory Data Analysis*. Addison-Wesley series in behavioral science. Addison-Wesley Publishing Company, 1977. ISBN 9780201076165.

G. B. Valsecchi, T. J. Jopek, and C. Froeschle. Meteoroid stream identification: a new approach - I. Theory. MNRAS, 304:743–750, April 1999. doi: 10.1046/j.1365-8711.1999.02264.x.

J.T. Vanderplas, A.J. Connolly, Ž. Ivezić, and A. Gray. Introduction to astroml: Machine learning for astrophysics. In *Conference on Intelligent Data Understanding (CIDU)*, pages 47 –54, oct. 2012. doi: 10.1109/CIDU.2012.6382200.

W.N. Venables and Brian D. Ripley. *Modern Applied Statistics with S (Statistics and Computing)*. Springer, 2003. ISBN 978-0-387-21706-2.

A. J. Verbiscer, M. F. Skrutskie, and D. P. Hamilton. SaturnÂt's largest ring. Nature, 461: 1098–1100, October 2009. doi: 10.1038/nature08515.

R. J. Weryk, M. D. Campbell-Brown, P. A. Wiegert, P. G. Brown, Z. Krzeminski, and R. Musci. The Canadian Automated Meteor Observatory (CAMO): System overview. Icarus, 225:614–622, July 2013. doi: 10.1016/j.icarus.2013.04.025.

F. L. Whipple. Photographic meteor orbits and their distribution in space. AJ, 59:201, July 1954. doi: 10.1086/106998.

D. T. Young, J.-J. Berthelier, M. Blanc, J. L. Burch, S. Bolton, A. J. Coates, F. J. Crary, R. Goldstein, M. Grande, T. W. Hill, R. E. Johnson, R. A. Baragiola, V. Kelha, D. J. McComas, K. Mursula, E. C. Sittler, K. R. Svenes, K. Szegö, P. Tanskanen, M. F. Thomsen, S. Bakshi, B. L. Barraclough, Z. Bebesi, D. Delapp, M. W. Dunlop, J. T. Gosling, J. D. Furman, L. K. Gilbert, D. Glenn, C. Holmlund, J.-M. Illiano, G. R. Lewis, D. R. Linder, S. Maurice, H. J. McAndrews, B. T. Narheim, E. Pallier, D. Reisenfeld, A. M. Rymer, H. T. Smith, R. L. Tokar, J. Vilppola, and C. Zinsmeyer. Composition and Dynamics of Plasma in Saturn's Magnetosphere. *Science*, 307:1262–1266, February 2005. doi: 10.1126/science.1106151.

Appendices

A. Orbit Dynamics

A.1. Reference Frames

Listing A.1 shows the frame kernel, that defines the Saturn-centric non-rotating coordinate system (SCNRC), for SPICE. The frame name is *SATURNJ2000* and the corresponding ID is defined in line 2 and 5. The first block of the frame definition sets miscellaneous parameters and conventions. The center is Saturn (line 6) and is coordinate system is defined with respect to J2000 (line 7).

Line 9 and 10 show, that the kernel is defined by using a parameterised form, based on two defined vectors. These vectors are then defined in the last two code blocks.

Before, a fixed date is set, where the reference frame refers to, since it is a fixed / intertial frame . The date (line 13) is the standard J2000 date and time.

From line 15 to 21 is the primary axis defined. Here it is the z-axis in positive direction (line 15). As mentioned in the main text of this work the z-axis is aligned with the rotational axis of Saturn. With respect to the J2000 system (line 17), the right ascension and declination values of the axis pointing are set in line 20 and 21, respectively. The dimension of the values is degrees (line 19).

The secondary axis is the x-axis and points as close as possbile towards the vernal equinox (line 23 to 27). However, due to the set primary axis, the x-axis is pointing finally on the vernal equinox corresponding meridian line. The axis is given in J2000 (line 25) as a rectangular pointing vector (line 26). By definition, the vernal equinox has the coordinates $(1, 0, 0)^{\mathrm{T}}$ (line 27).

```
1  \begindata
2     FRAME_SATURNJ2000              =  1400699
3     FRAME_1400699_NAME             = 'SATURNJ2000'
4     FRAME_1400699_CLASS            =  5
5     FRAME_1400699_CLASS_ID         =  1400699
6     FRAME_1400699_CENTER           =  699
7     FRAME_1400699_RELATIVE         = 'J2000'
8
9     FRAME_1400699_DEF_STYLE        = 'PARAMETERIZED'
10    FRAME_1400699_FAMILY           = 'TWO-VECTOR'
11
12    FRAME_1400699_ROTATION_STATE   = 'INERTIAL'
13    FRAME_1400699_FREEZE_EPOCH     =  @2000-01-01/12:00:00
14
15    FRAME_1400699_PRI_AXIS         = '+Z'
16    FRAME_1400699_PRI_VECTOR_DEF   = 'CONSTANT'
```

```
   FRAME_1400699_PRI_FRAME       = 'J2000'
   FRAME_1400699_PRI_SPEC        = 'RA/DEC'
   FRAME_1400699_PRI_UNITS       = 'DEGREES'
   FRAME_1400699_PRI_RA          = 40.581871
   FRAME_1400699_PRI_DEC         = +83.537722

   FRAME_1400699_SEC_AXIS        = '+X'
   FRAME_1400699_SEC_VECTOR_DEF  = 'CONSTANT'
   FRAME_1400699_SEC_FRAME       = 'J2000'
   FRAME_1400699_SEC_SPEC        = 'RECTANGULAR'
   FRAME_1400699_SEC_VECTOR      = ( 1., 0., 0. )
```

Listing A.1: SPICE frame kernel of the SCNRC

Lisiting A.2 lists the content of the Saturn-centric rotating coordinate system (SCRC) frame. As before, the first two blocks define the name, ID, center of the frame etc. Since the SCRC depends on the position of Saturn, no intertial time is set. The x-axis is the primary axis in this case and is defined in the following block. The vector points from Saturn to the Sun. The last block sets the z-axis points to the ecliptic North Pole. Listing A.3 is a similar reference frame for the Earth.

```
\begindata
   FRAME_SATURN_SUN              =  1401699
   FRAME_1401699_NAME            = 'SATURN_SUN'
   FRAME_1401699_CLASS           =  5
   FRAME_1401699_CLASS_ID        =  1401699
   FRAME_1401699_CENTER          =  699
   FRAME_1401699_RELATIVE        = 'ECLIPJ2000'

   FRAME_1401699_DEF_STYLE       = 'PARAMETERIZED'
   FRAME_1401699_FAMILY          = 'TWO-VECTOR'

   FRAME_1401699_PRI_AXIS        = '+X'
   FRAME_1401699_PRI_VECTOR_DEF  = 'OBSERVER_TARGET_POSITION'
   FRAME_1401699_PRI_OBSERVER    = 'SATURN'
   FRAME_1401699_PRI_TARGET      = 'SUN'
   FRAME_1401699_PRI_ABCORR      = 'NONE'

   FRAME_1401699_SEC_AXIS        = '+Z'
   FRAME_1401699_SEC_VECTOR_DEF  = 'CONSTANT'
   FRAME_1401699_SEC_FRAME       = 'ECLIPJ2000'
   FRAME_1401699_SEC_SPEC        = 'RECTANGULAR'
   FRAME_1401699_SEC_VECTOR      = (0,
                                    0,
                                    1)
```

Listing A.2: SPICE frame kernel of the SCRC

```
\begindata
   FRAME_EARTH_SUN               =  1400310
   FRAME_1400310_NAME            = 'EARTH_SUN'
   FRAME_1400310_CLASS           =  5
   FRAME_1400310_CLASS_ID        =  1400310
   FRAME_1400310_CENTER          =  3
   FRAME_1400310_RELATIVE        = 'ECLIPJ2000'

   FRAME_1400310_DEF_STYLE       = 'PARAMETERIZED'
```

```
10    FRAME_1400310_FAMILY          = 'TWO-VECTOR'
11
12    FRAME_1400310_PRI_AXIS        = '+X'
13    FRAME_1400310_PRI_VECTOR_DEF  = 'OBSERVER_TARGET_POSITION'
14    FRAME_1400310_PRI_OBSERVER    = 3
15    FRAME_1400310_PRI_TARGET      = 10
16    FRAME_1400310_PRI_ABCORR      = 'NONE'
17
18    FRAME_1400310_SEC_AXIS        = '+Z'
19    FRAME_1400310_SEC_VECTOR_DEF  = 'CONSTANT'
20    FRAME_1400310_SEC_FRAME       = 'ECLIPJ2000'
21    FRAME_1400310_SEC_SPEC        = 'RECTANGULAR'
22    FRAME_1400310_SEC_VECTOR      = ( 0,
23                                      0,
24                                      1 )
```

Listing A.3: SPICE frame kernel of the ECRC

B. Canary Island Long-Baseline Observatory (CILBO)

In this chapter the content of the CILBO orbit database is shown (table B.2). For better readability, and to compress the size of the table, several entries were replaced with a placeholder called *SUFFIX*. The correspond list of suffixes is shown in table B.1. E.g.: the entry X01 called *appmag_SUFFIX* is in the database: *appmag_mean*, *appmag_mean_std*, *appmag_median* and *appmag_median_mad*.

Table B.1.: List of all suffixes for parameters listed in table B.2.

Suffix	Description
mean	Mean
mean_std	Standard Deviation
median	Median
median_mad	Median Absolute Deviation

Table B.2.: Entries of the CILBO MOTS orbit database table *icc*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity. *SUFFIX* is a placeholder for the ending terms shown in table B.1 for better readability.

CN	Column Name	Data Type	Description	Unit
001	meteor_id	INT	Individual ID of a meteor entry; format YYYYMMDDhhmmss (YYYY: Year, MM: Month, DD: Day, hh: hour, mm: minute, ss: second)	—
002	cam1_frames_bias	INT	Number of frames from Intensified CCD Camera (ICC)7 (all frames)	—
003	cam1_frames	INT	Number of frames from ICC7 (Random Sample Consensus (RANSAC) corrected)	DN
004	cam2_frames_bias	INT	Number of frames from ICC9 (all frames)	—
005	cam2_frames	INT	Number of frames from ICC9 (RANSAC corrected)	—
006	et4spice1	INT	ICC7 detection time (first frame) in ephemeris time	s
007	utc4spice1	TEXT	ICC7 detection time (first frame) in Coordinated Universal Time (UTC)	Date
008	et4spice2	INT	ICC9 detection time (first frame) in ephemeris time	s
009	utc4spice2	TEXT	ICC9 detection time (first frame) in UTC	Date
010	seeds	INT	Number of Monte-Carlo (MC) seeds	—
011	vel_type	TEXT	Velocity determination type	None
X01	appmag_SUFFIX	REAL	Apparent magnitude	mag
012	measure_acc_cam1	REAL	Accuracy (log-file from ICC7)	arcmin
013	measure_acc_cam2	REAL	Accuracy (log-file from ICC9)	arcmin
X02	fit_cam1_astrometry_dist_SUFFIX	REAL	Fit accuracy w.r.t. actual data points (ICC7)	m
X03	fit_cam2_astrometry_dist_SUFFIX	REAL	Fit accuracy w.r.t. actual data points (ICC9)	m

014	fit_cam1_astrometry_dist_max_median	REAL	Fit accuracy w.r.t. actual data points (maximum median, ICC7)	m
015	fit_cam1_astrometry_dist_max_median_mad	REAL	Fit accuracy w.r.t. actual data points (maximum Median Absolute Deviation (MAD), ICC7)	m
016	fit_cam2_astrometry_dist_max_median	REAL	Fit accuracy w.r.t. actual data points (maximum median, ICC9)	m
017	fit_cam2_astrometry_dist_max_median_mad	REAL	Fit accuracy w.r.t. actual data points (maximum MAD, ICC9)	m
018	fit_cam1_astrometry_dist_bias_max	REAL	Fit accuracy w.r.t. actual data points (maximum from biased data set, ICC7)	m
019	fit_cam1_astrometry_dist_unbiased_max	REAL	Fit accuracy w.r.t. actual data points (maximum from unbiased data set, ICC7)	m
020	fit_cam2_astrometry_dist_bias_max	REAL	Fit accuracy w.r.t. actual data points (maximum from biased data set, ICC9)	m
021	fit_cam2_astrometry_dist_unbiased_max	REAL	Fit accuracy w.r.t. actual data points (maximum from unbiased data set, ICC9)	m
X04	vel_SUFFIX	REAL	Entry velocity	km/s
X05	altitude_SUFFIX	REAL	Altitude of meteor	km
X06	zenith_distance_SUFFIX	REAL	Angular distance between radiant and zenith	deg
X07	angular_distance_SUFFIX	REAL	Angular distance between radiant and boresight	deg
X08	radiant_ra_SUFFIX	REAL	Radiant (right ascension)	deg
X09	radiant_dec_SUFFIX	REAL	Radiant (declination)	deg
022	radiant_cov_ra_dec	REAL	Covariance matrix (ra, dec)	MultiDim.
X10	radiant_ecsc_long_SUFFIX	REAL	Radiant (longitude in Earth-centric rotating coordinate system (ECRC) at Earth's Sphere Of Influence (SOI))	deg
X11	radiant_ecsc_lat_SUFFIX	REAL	Radiant (latitude in ECRC at Earth's SOI)	deg
023	radiant_cov_ecsc_long_lat	TEXT	Covariance Matrix; ECRC (long, lat)	MultiDim.

X12	radiant_eclip_long_SUFFIX	REAL	Radiant (longitude in ECLIPJ2000 at Earth's SOI)	deg
X13	radiant_eclip_lat_SUFFIX	REAL	Radiant (latitude in ECLIPJ2000 at Earth's SOI)	deg
024	radiant_cov_eclip_long_lat	TEXT	Covariance Matrix; ECLIPJ2000 (long, lat)	MultiDim.
X14	meteor_appmax_cam_ distance_SUFFIX	REAL	Distance to meteor at meteor's maximum brightness	km
X15	abs_mag_SUFFIX	REAL	Absolute magnitude	mag
X16	q_earth_SUFFIX	REAL	Perigee in Earth-Centric Coordinate System (ECCS)	ER
X17	e_earth_SUFFIX	REAL	Eccentricity in ECCS	—
X18	i_earth_SUFFIX	REAL	Inclination in ECCS	deg
X19	lnode_earth_SUFFIX	REAL	Longitude of ascending node in ECCS	deg
X20	argp_earth_SUFFIX	REAL	Argument of perigee in ECCS	deg
X21	mzero_earth_SUFFIX	REAL	Mean anomaly at *tzero* in ECCS	deg
025	tzero_earth	REAL	Epoch in ECCS	s
026	conics_earth_cov	TEXT	Covariance matrix (order: see above, from *q* to *mzero*)	MultiDim.
X22	q_helio_SUFFIX	REAL	Perihelion in ECLIPJ2000	Astronomical Unit (AU)
X23	e_helio_SUFFIX	REAL	Eccentricity in ECLIPJ2000	—
X24	i_helio_SUFFIX	REAL	Inclination in ECLIPJ2000	deg
X25	lnode_helio_SUFFIX	REAL	Longitude of ascending node in ECLIPJ2000	deg
X26	argp_helio_SUFFIX	REAL	Argument of perihelion in ECLIPJ2000	deg
X27	mzero_helio_SUFFIX	REAL	Mean anomaly at *tzero* in ECLIPJ2000	deg
027	tzero_helio	REAL	Epoch in ECLIPJ2000	s
028	conics_helio_cov	TEXT	Covariance matrix (order: see above, from *q* to *mzero*)	MultiDim.
029	earth_bound_orbit_prob	REAL	Probability of an Earth-bound particle	—
X28	tisserand_SUFFIX	REAL	Tisserand parameter w.r.t. to Jupiter	—

C. Cosmic-Dust-Analyzer (CDA)

C.1. Sensitive area / angular Field-of-view (FOV)

The following list shows the content of the file *sensitive_area_with_w2.cda.dat* (limited until last entry for the Impact Ionisation Detector (IID)), generated by Srama [priv. comm.]. The content of the file shows the sensitive area of CDA vs. the impact angle in degrees. The step size is 0.25° and 0.0° refers to an anti-parallel impact direction with respect to the boresight. The transmission of four grids has been taken into account and are shown in the second header line (first 3 lines are the header). The forth line shows the description of the columns. The first entry is the impact angle, the second and third column list the sensitive area of IID and Chemical Analyzer Target (CAT) in m^2, respectively. The fourth column is the total sensitive area that is not considering the fifth column where the sensitive area of the wall is shown. The two last entries are the integrated sensitive areas for the main target IID and CAT in m^2. For the study, only the IID and CAT have been considered.

Data C.1.1.: Simulated sensitive area data of the CDA. The first columns lists shows the impact angle in 1 degrees steps. The second and third column list the sensitive area in square meters for the smaller and larger target, respectively. The fourth columns shows the total area of both targets. The following column shows the area of the inner Wall and the last two entries are the integrated area for the small and large target. Note: The original data set has a resolution of 0.25 degrees and has been reduced here for better readability.

```
Data produced by program sensitive_area_w.pro, R. Srama
Mon Jan 11 13:20:46 2016 with QP transmission 0.95^4 and CA-grid 68% 0.25 deg step
CDA sensitive area in m^2 including QP grid transmission
angle      IIT            CAT            total           innerWall    integralCAT    integralIIT
0.00000    0.0881859      0.00663086     0.0948167      0.00522424    0.0881859      0.00663086
1.00000    0.0860244      0.00659098     0.0926154      0.00525194    0.435554       0.0330646
2.00000    0.0838618      0.00654224     0.0904040      0.00525519    0.774239       0.0593078
3.00000    0.0816514      0.00649683     0.0881482      0.00582698    1.10416        0.0853660
4.00000    0.0794122      0.00644809     0.0858602      0.00657306    1.42518        0.111237
5.00000    0.0771774      0.00640821     0.0835856      0.00735499    1.73725        0.136923
6.00000    0.0749060      0.00636058     0.0812666      0.00818090    2.04023        0.162434
7.00000    0.0725914      0.00632070     0.0789121      0.00901984    2.33406        0.187784
8.00000    0.0702639      0.00630630     0.0765702      0.00986530    2.61863        0.212997
9.00000    0.0678010      0.00637830     0.0741793      0.0107205     2.89354        0.238362
10.0000    0.0652576      0.00650901     0.0717666      0.0115660     3.15836        0.264200
11.0000    0.0626859      0.00665302     0.0693389      0.0124212     3.41296        0.290589
12.0000    0.0601420      0.00677708     0.0669190      0.0132699     3.65735        0.317511
13.0000    0.0575659      0.00691776     0.0644837      0.0141235     3.89148        0.344950
14.0000    0.0549869      0.00704404     0.0620309      0.0149527     4.11530        0.372932
15.0000    0.0524189      0.00716922     0.0595881      0.0157705     4.32876        0.401470
```

16.0000	0.0501689	0.00699530	0.0571642	0.0166176	4.53271	0.429782
17.0000	0.0481317	0.00665634	0.0547880	0.0174435	4.72827	0.456952
18.0000	0.0461652	0.00624316	0.0524084	0.0182563	4.91587	0.482563
19.0000	0.0442408	0.00579342	0.0500342	0.0190757	5.09574	0.506418
20.0000	0.0423222	0.00535697	0.0476792	0.0198886	5.26787	0.528514
21.0000	0.0404455	0.00486293	0.0453085	0.0206819	5.43249	0.548710
22.0000	0.0386622	0.00432457	0.0429868	0.0214688	5.58978	0.566833
23.0000	0.0369437	0.00371421	0.0406580	0.0222849	5.74013	0.582611
24.0000	0.0352551	0.00311050	0.0383656	0.0230391	5.88368	0.595956
25.0000	0.0335623	0.00250900	0.0360713	0.0238194	6.02044	0.606886
26.0000	0.0318358	0.00193963	0.0337754	0.0245899	6.15037	0.615498
27.0000	0.0301130	0.00141235	0.0315254	0.0253409	6.27340	0.621933
28.0000	0.0283430	0.000926061	0.0292691	0.0260886	6.38945	0.626354
29.0000	0.0265562	0.000504016	0.0270602	0.0268347	6.49837	0.628974
30.0000	0.0246543	0.000170590	0.0248249	0.0275678	6.59984	0.630121
31.0000	0.0226319	0.000000000	0.0226319	0.0282748	6.69353	0.630292
32.0000	0.0205500	0.000000000	0.0205500	0.0289883	6.77881	0.630292
33.0000	0.0186978	0.000000000	0.0186978	0.0296871	6.85633	0.630292
34.0000	0.0169385	0.000000000	0.0169385	0.0303599	6.92670	0.630292
35.0000	0.0152590	0.000000000	0.0152590	0.0310359	6.99020	0.630292
36.0000	0.0136479	0.000000000	0.0136479	0.0316989	7.04716	0.630292
37.0000	0.0120498	0.000000000	0.0120498	0.0323522	7.09777	0.630292
38.0000	0.0104713	0.000000000	0.0104713	0.0329859	7.14201	0.630292
39.0000	0.00888138	0.000000000	0.00888138	0.0336049	7.17992	0.630292
40.0000	0.00733381	0.000000000	0.00733381	0.0342093	7.21155	0.630292
41.0000	0.00584978	0.000000000	0.00584978	0.0347859	7.23716	0.630292
42.0000	0.00447978	0.000000000	0.00447978	0.0352681	7.25711	0.630292
43.0000	0.00322707	0.000000000	0.00322707	0.0356591	7.27186	0.630292
44.0000	0.00221383	0.000000000	0.00221383	0.0359556	7.28213	0.630292
45.0000	0.00159480	0.000000000	0.00159480	0.0361119	7.28940	0.630292
46.0000	0.000957859	0.000000000	0.000957859	0.0361347	7.29412	0.630292
47.0000	0.000390963	0.000000000	0.000390963	0.0359214	7.29659	0.630292
48.0000	1.14031e-05	0.000000000	1.14031e-05	0.0355939	7.29707	0.630292

C.2. Dust Orbit Computation Code (DOCC) test cases

The following figures show miscellaneous orbital element distributions from the DOCC test case. Each figure (with 4 sub-figure) summarizes differently set re-sampling sizes, to visualize the resulting distributions and statistical improvements.

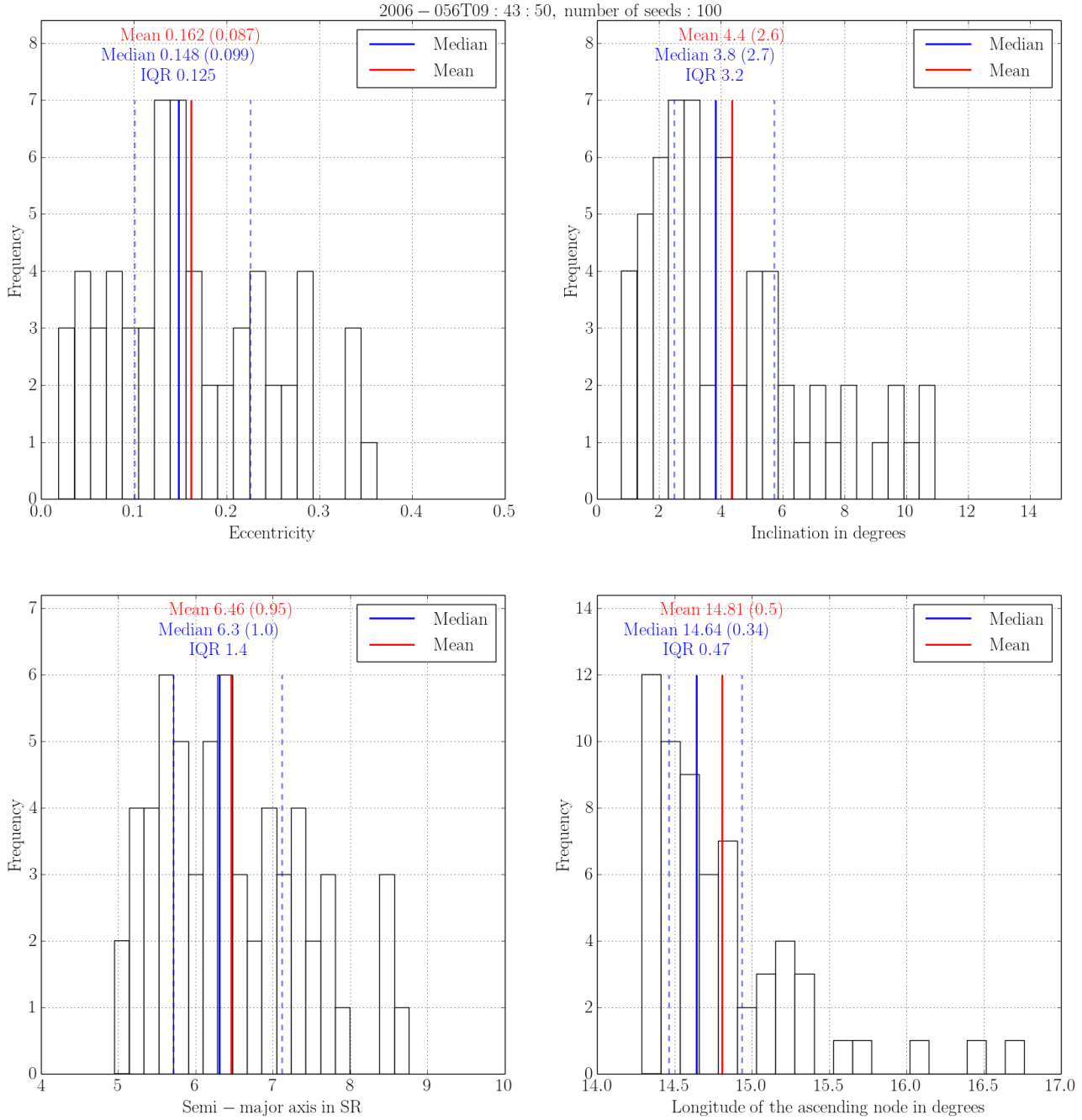


Figure C.1.: Determined orbital elements of a Chemical Analyzer Target impact at 2006-056T09:43:50. Left top panel to bottom right: eccentricity, inclination, semi-major axis, longitude of the ascending node in Saturn-centric non-rotating coordinate system. Results based on a 100 seeds Monte-Carlo simulation. The solid red and blue lines indicate the mean and median respectively. The blue dashed lines show the corresponding Inter Quartile Range (IQR) boundaries. Above the lines the results are listed (top to bottom): mean and standard deviation, median and Median Absolute Deviation (MAD), Inter Quartile Range (IQR).

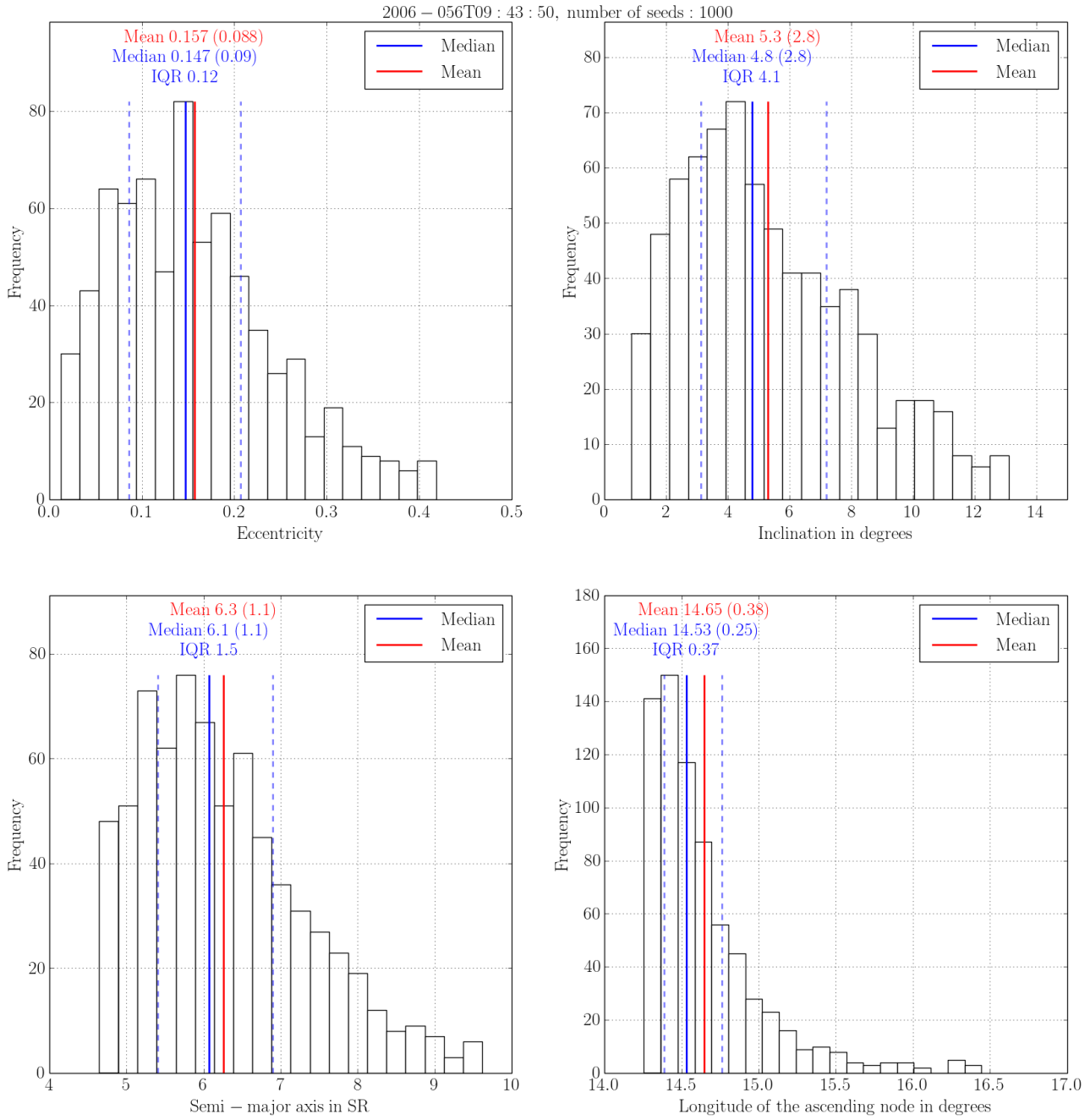


Figure C.2.: Determined orbital elements of a Chemical Analyzer Target impact at 2006-056T09:43:50. Left top panel to bottom right: eccentricity, inclination, semi-major axis, longitude of the ascending node in Saturn-centric non-rotating coordinate system. Results based on a 1,000 seeds Monte-Carlo simulation. The solid red and blue lines indicate the mean and median respectively. The blue dashed lines show the corresponding Inter Quartile Range (IQR) boundaries. Above the lines the results are listed (top to bottom): mean and standard deviation, median and Median Absolute Deviation (MAD), Inter Quartile Range (IQR).

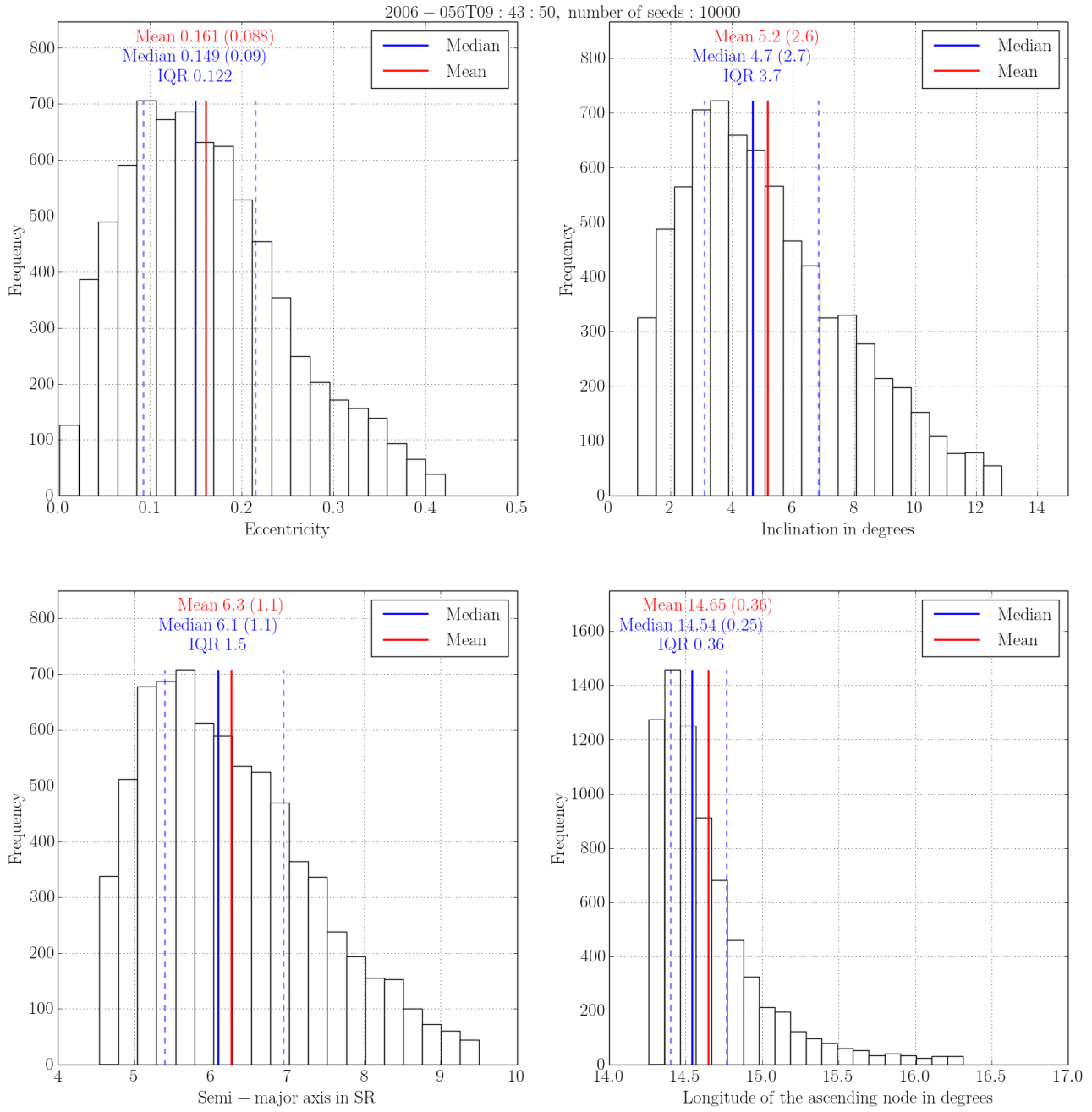


Figure C.3.: Determined orbital elements of a Chemical Analyzer Target impact at 2006-056T09:43:50. Left top panel to bottom right: eccentricity, inclination, semi-major axis, longitude of the ascending node in Saturn-centric non-rotating coordinate system. Results based on a 10,000 seeds Monte-Carlo simulation. The solid red and blue lines indicate the mean and median respectively. The blue dashed lines show the corresponding Inter Quartile Range (IQR) boundaries. Above the lines the results are listed (top to bottom): mean and standard deviation, median and Median Absolute Deviation (MAD), Inter Quartile Range (IQR).

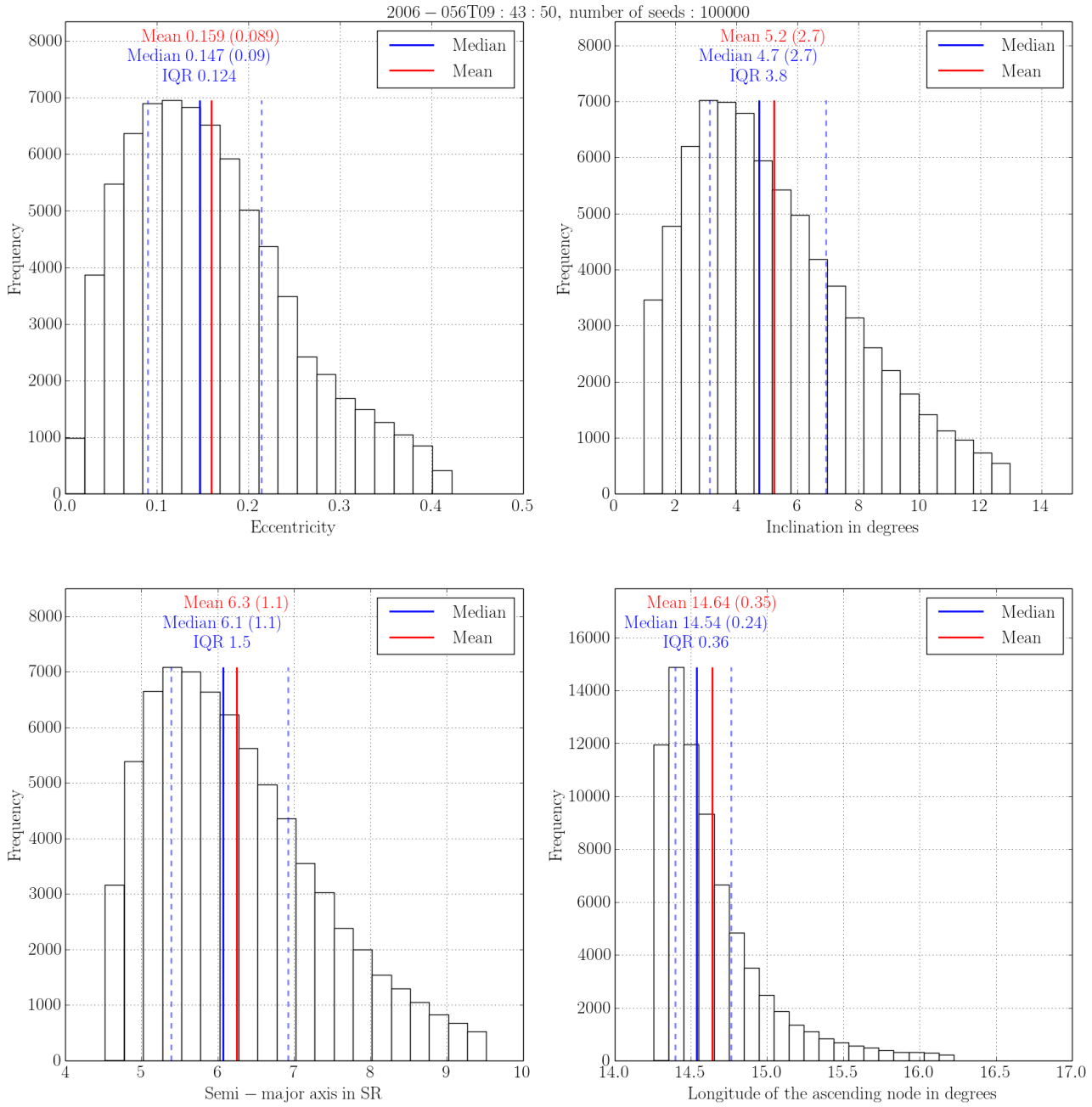


Figure C.4.: Determined orbital elements of a Chemical Analyzer Target impact at 2006-056T09:43:50. Left top panel to bottom right: eccentricity, inclination, semi-major axis, longitude of the ascending node in Saturn-centric non-rotating coordinate system. Results based on a 100,000 seeds Monte-Carlo simulation. The solid red and blue lines indicate the mean and median respectively. The blue dashed lines show the corresponding Inter Quartile Range (IQR) boundaries. Above the lines the results are listed (top to bottom): mean and standard deviation, median and Median Absolute Deviation (MAD), Inter Quartile Range (IQR).

Table C.1.: Descriptive parameters of the eccentricity results for a seed size of 100, 1k, 10k and 100k.

Seed	Median e	MAD e	IQR e
100	0.148	0.099	0.125
1k	0.147	0.090	0.120
10k	0.149	0.090	0.122

100k	0.147	0.090	0.124

Table C.2.: Descriptive parameters of the inclination results for a seed size of 100, 1k, 10k and 100k.

Seed	Median i / deg	MAD i / deg	IQR i / deg
100	3.8	2.7	3.2
1k	4.8	2.8	4.1
10k	4.7	2.7	3.7
100k	4.7	2.7	3.8

Table C.3.: Descriptive parameters of the semi-major axis (in SR) results for a seed size of 100, 1k, 10k and 100k.

Seed	Median a / SR	MAD a / SR	IQR a / SR
100	6.3	1.0	1.4
1k	6.1	1.1	1.5
10k	6.1	1.1	1.5
100k	6.1	1.1	1.5

Table C.4.: Descriptive parameters of the longitude of the ascending node results for a seed size of 100, 1k, 10k and 100k.

Seed	Median lnode / deg	MAD lnode / deg	IQR lnode / deg
100	14.64	0.34	0.47
1k	14.53	0.25	0.37
10k	14.54	0.25	0.36
100k	14.54	0.24	0.36

C.3. DOCC input file

DOCC requires input files with a certain structure and content. The following example shows an IID input file. For better readability the content was separated in 3 listings. The first two header lines are ignored by the software, however, the order of the content needs to be considered. The entry names correspond to the structure explained in Albin [2014].

Listing C.1: Part1 of an IID DOCC input file

```
SCLK          SCET                  cda_counter  cda_qi_tr  cda_qi_amp    cda_nr_mp_peaks
------------  --------------------  -----------  ---------  ------------  ---------------
1457775003    2004-072T09:06:50     10.0         8.8330E-6  6.10700E-15   0.0
```

Listing C.2: Part2 of an IID DOCC input file

```
cda_bs_x  cda_bs_y  cda_bs_z  cda_sun  cda_isd  cda_idp cda_dust_ram_pro_saturn
--------  --------  --------  -------  -------  ------- -------------------------------
0.282093  -0.86302  0.419072  47.1650  61.9526  101.679 109.34563
```

Listing C.3: Part3 of an IID DOCC input file

```
cda_dust_ram_retro_saturn  cda_saturn  cda_ring
-------------------------  ----------  --------
93.6742                    101.8345    22.1687
```

C.4. DOCC database content

This section describes the CDA orbit database tables in each subsection. Note: The Spacecraft Event Time (SCET) is saved as an integer: YYYYMMDDhhmmss (YYYY: Year, MM: Month, DD: Day, hh: Hours, mm: Minutes, ss: seconds). The covariance matrices for the orbits have the order: q, e, a, i, lnode, argp, mzero, tzero and the order for the radiant matrices are: lon, lat, vel (check the corresponding tables for explanations).

The current version of the database can be downloaded from:
sftp://USERNAME@idust.irs.uni-stuttgart.de: albin/CDA/DOCC2/cda_ orbit.db

Spacecraft Clock (SCLK) is the primary key of the *main* table and foreign key for all other corresponding tables. 1 Saturn Radius / Radii (SR) correponds to the equatirial radius of Saturn: 60,278 km.

C.4.1. Table: main

Table C.5 shows the content of table the prime table *main*. The table saves some geometries like the distance of Cassini to Saturn and the Sun as well as miscellaneous angles between CDA's boresight and e.g. the Dust-Ram.

Table C.5.: Entries of the orbit database table *main*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity.

CN	Column Name	Data Type	Description	Unit
01	SCLK	INTEGER	Spacecraft clock	DN
02	SCET	INTEGER	Spacecraft event time	DN
03	et	FLOAT64	ephemeris time	s
04	cassini_saturn_dist	FLOAT64	Distance Cassini - Saturn	SR

05	cassini_saturn_z_dist	FLOAT64	Distance Cassini - Ring Plane	SR
06	cassini_sun_dist	FLOAT64	Distance Cassini - Sun	SR
07	cda_sun	FLOAT64	Angular distance CDA - Sun	deg
08	cda_isd	FLOAT64	Angular distance CDA - Interplanetary Dust Population (IDP) direction	deg
09	cda_idp	FLOAT64	Angular distance CDA - Interstellar Direction (ISD)	deg
10	cda_dust_ram	FLOAT64	Angular distance CDA - dust ram	deg
11	cda_anti_dust_ram	FLOAT64	Angular distance CDA - anti dust ram	deg
12	cda_saturn	FLOAT64	Angular distance CDA - Saturn	deg
13	cda_ring	FLOAT64	Angular distance CDA - Ring plane	deg

C.4.2. Table: event

Table C.6 shows the content of the table *event* and saves some event data of an impact event. This table summarizes some instrument determined parameters like the velocity, mass, amplitude and signal rise time. All possible solution probabilities are also listed.

Table C.6.: Entries of the orbit database table *event*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity.

CN	Column Name	Data Type	Description	Unit
01	SCLK	INTEGER	Spacecraft clock	DN
02	SCET	INTEGER	Spacecraft event time	DN
03	et	FLOAT64	ephemeris time	DN
04	target_id	INTEGER	For IID impact: 1, for CAT impact: 2	DN
05	cda_counter	INTEGER	Counter class number of the impact	DN
06	event_vel	FLOAT64	Impact velocity based on Srama [2009] eq. 3.14 and 3.29 for IID and CAT, respectively	km/s

07	event_mass	FLOAT64	Mass of the dust particle, based on Srama [2009] eq. 3.28 and 3.33 for IID and CAT, respectively	kg
08	event_tr	FLOAT64	Signal rise time of QI (IID) or QC (CAT) signal	μs
09	event_amp	FLOAT64	Ampltiude of the QI (IID) or QC (CAT) signal	C
10	event_nr_peaks	INTEGER	Number of QM peaks (if available)	DN
11	isd_prob	FLOAT64	ISD probability	DN
12	helio_prob	FLOAT64	IDP probability	DN
13	saturn_prob	FLOAT64	Saturn-bound particle probability	DN
14	moon_prob	FLOAT64	Moon-bound particle probability	DN
15	moon_id	INTEGER	Corresponding moon ID (see table 5.3)	DN
16	moon_ini_prob	FLOAT64	Moon-escape probability	DN
17	moon_source_prob	FLOAT64	Moon surface source probability	DN
18	saturn_inbound_prob	FLOAT64	Saturn-inbound probability	DN
19	saturn_outbound_prob	FLOAT64	Saturn-outbound probability	DN
20	moon_inbound_prob	FLOAT64	Moon-inbound probability	DN
21	moon_outbound_prob	FLOAT64	Moon-outbound probability	DN

C.4.3. Table: monte_carlo_summary

Table C.7 shows the database table monte_carlo_summary that summarizes some meta parameters.

Table C.7.: Entries of the orbit database table *monte_carlo_summary*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity.

CN	Column Name	Data Type	Description	Unit
01	seeds	INTEGER	Number of seeds	DN
02	proc	INTEGER	Number of used central processing units	
03	cs	INTEGER	Chuncksize	

04	docc_version	INTEGER	Version number of the code	DN
05	user_id	STRING	User string	str
06	computer_id	STRING	Computer string	str

C.4.4. Table: moon_id

Table C.8 shows the content of table *moon_id* that lists some properties of 8 Saturn moons that are used in the orbit database. The corresponding data are listed in table 5.3.

Table C.8.: Entries of the orbit database table *moon_id*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity. *MOON* is a placeholder for the moons shown in table 5.3.

CN	Column Name	Data Type	Description	Unit
X1	MOON_id	INTEGER	Navigation and Ancillary Information Facility MOON ID	DN
X2	MOON_mr	INTEGER	Mean radiu of MOON	km
X3	MOON_mass	FLOAT64	Mass of MOON	kg
X4	MOON_hsr	INTEGER	Hill sphere radius of MOON	km
X5	MOON_a	FLOAT64	Semi-major axis of MOON	SR

C.4.5. Astrodynamical parameters tables

This section shows the orbit database tables that contain the astrodynamical data. E.g. table *isd* lists all properties of particles that have an eccentricity larger than 1 in Helio-centric ecliptic coordinates (HCEC) with a probability larger than 0.

A Monte-Carlo simulation returns a set of possible solutions. Thus, for each parameter 11 statistical quantities are saved in the database that allow a future sophisticated scientific analysis. The name of each quantity is the suffix for each parameter shown in the following subsections. These suffixes are shown in tables C.9. E.g.: to obtain the upper IQR border inclination of a saturn bound particle, a user needs to extract the entry *i_saturn_iqr_u* from the table *saturn*. All table entries can be obtained in the SQLite3 shell with the command *.schema TABLENAME*.

Table C.9.: List of all suffixes for parameters listed in the following sub-sections. Note: Each word is separated by an underline. E.g. the entry *a_saturn* from the table *saturn* becomes for the upper confidence border: *a_saturn_conf_u*.

Suffix	Description
median	Median
mad	Median Absolute Deviation
iqr_l	Lower border of the IQR
iqr_u	Upper border of the IQR

iqr	Width of the IQR
notch_l	Lower border of the notch
notch_u	Upper border of the notch
notch	Width of the notch
conf_l	Lower border of the confidence interval
conf_u	Upper border of the confidence interval
conf	Width of the confidence interval

Table: isd

Table C.10 shows the content of table *isd*. The table contains only orbital elements of ISD particles.

Table C.10.: Entries of the orbit database table *isd*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity. Orbital elements are given with respect to EclipJ2000 (HCEC).

CN	Column Name	Data Type	Description	Unit
001	SCLK	INTEGER	Spacecraft clock	DN
002	SCET	INTEGER	Spacecraft event time	DN
003	et	FLOAT64	ephemeris time	DN
004	isd_prob	FLOAT64	ISD probability	DN
X01	q_isd_SUFFIX	FLOAT64	Perihelion	SR
X02	e_isd_SUFFIX	FLOAT64	Eccentricity	DN
X03	a_isd_SUFFIX	FLOAT64	Semi-major axis	SR
X04	i_isd_SUFFIX	FLOAT64	Inclination	deg
X05	lnode_isd_SUFFIX	FLOAT64	Long. of asc. node	deg
X06	argp_isd_SUFFIX	FLOAT64	Argument of periapsis	deg
X07	mzero_isd_SUFFIX	FLOAT64	True anomaly at epoch	deg
X08	tzero_isd_SUFFIX	FLOAT64	Corresponding epoch	s
005	conics_isd_cov	BLOB	Covariance matrix of the orbital elements	misc
X09	isd_eclip_lon_SUFFIX	FLOAT64	Radiant ecliptic longitude	deg
X10	isd_eclip_lat_SUFFIX	FLOAT64	Radiant ecliptic latitude	deg
X11	isd_scrc_lon_SUFFIX	FLOAT64	Radiant SCRC longitude	deg
X12	isd_scrc_lat_SUFFIX	FLOAT64	Radiant SCRC latitude	deg
X13	isd_eclip_vel_SUFFIX	FLOAT64	SOI intersection ecliptic velocity	km/s
X14	isd_scrc_vel_SUFFIX	FLOAT64	SOI intersection SCRC velocity	km/s

006	radiant_isd_eclip_cov	BLOB	Covariance matrix of ecliptic radiant information	misc
007	radiant_isd_scrc_cov	BLOB	Covariance matrix of SCRC radiant information	misc

Table: helio

Table C.11 shows the content of table *helio*. Orbital elements in this table correspond to a Sun-bound orbit (IDP).

Table C.11.: Entries of the orbit database table *helio*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity. Orbital elements are given with respect to EclipJ2000 (HCEC).

CN	Column Name	Data Type	Description	Unit
001	SCLK	INTEGER	Spacecraft clock	DN
002	SCET	INTEGER	Spacecraft event time	DN
003	et	FLOAT64	ephemeris time	DN
004	helio_prob	FLOAT64	IDP probability	DN
X01	q_helio_SUFFIX	FLOAT64	Perihelion	SR
X02	e_helio_SUFFIX	FLOAT64	Eccentricity	DN
X03	a_helio_SUFFIX	FLOAT64	Semi-major axis	SR
X04	i_helio_SUFFIX	FLOAT64	Inclination	deg
X05	lnode_helio_SUFFIX	FLOAT64	Long. of asc. node	deg
X06	argp_helio_SUFFIX	FLOAT64	Argument of periapsis	deg
X07	mzero_helio_SUFFIX	FLOAT64	True anomaly at epoch	deg
X08	tzero_helio_SUFFIX	FLOAT64	Corresponding epoch	s
005	conics_helio_cov	BLOB	Covariance matrix of the orbital elements	misc
X09	helio_eclip_lon_SUFFIX	FLOAT64	Radiant ecliptic longitude	deg
X10	helio_eclip_lat_SUFFIX	FLOAT64	Radiant ecliptic latitude	deg
X11	helio_scrc_lon_SUFFIX	FLOAT64	Radiant SCRC longitude	deg
X12	helio_scrc_lat_SUFFIX	FLOAT64	Radiant SCRC latitude	deg
X13	helio_eclip_vel_SUFFIX	FLOAT64	SOI intersection ecliptic velocity	km/s
X14	helio_scrc_vel_SUFFIX	FLOAT64	SOI intersection SCRC velocity	km/s
006	radiant_helio_eclip_cov	BLOB	Covariance matrix of ecliptic radiant information	misc
007	radiant_helio_scrc_cov	BLOB	Covariance matrix of SCRC radiant information	misc

Table: saturn

Table C.12 shows the content of table *saturn*. It contains dust particles on a Saturn-bound orbit.

Table C.12.: Entries of the orbit database table *saturn*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity. *MOON* is a placeholder for the moons shown in table 5.3. Orbital elements are given with respect to SCNRC.

CN	Column Name	Data Type	Description	Unit
001	SCLK	INTEGER	Spacecraft clock	DN
002	SCET	INTEGER	Spacecraft event time	DN
003	et	FLOAT64	ephemeris time	DN
004	saturn_prob	FLOAT64	Saturn-bound probability	DN
X01	q_saturn_SUFFIX	FLOAT64	Perihelion	SR
X02	e_saturn_SUFFIX	FLOAT64	Eccentricity	DN
X03	a_saturn_SUFFIX	FLOAT64	Semi-major axis	SR
X04	i_saturn_SUFFIX	FLOAT64	Inclination	deg
X05	lnode_saturn_SUFFIX	FLOAT64	Long. of asc. node	deg
X06	lnode_rot_saturn_SUFFIX	FLOAT64	Long. of asc. node (Saturn_Sun)	deg
X07	argp_saturn_SUFFIX	FLOAT64	Argument of periapsis	deg
X08	mzero_saturn_SUFFIX	FLOAT64	True anomaly at epoch	deg
X09	tzero_saturn_SUFFIX	FLOAT64	Corresponding epoch	s
005	conics_saturn_cov	BLOB	Covariance matrix of the orbital elements	misc
X10	tisserand_MOON_SUFFIX	FLOAT64	Tisserand parameter with respect to MOON	DN

Table: moon

Table C.13 shows the content of table *moon*. It contains dust particles on a Moon-bound orbit.

Table C.13.: Entries of the orbit database table *moon*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity. Orbital elements are given with respect to MCNRC.

CN	Column Name	Data Type	Description	Unit
001	SCLK	INTEGER	Spacecraft clock	DN
002	SCET	INTEGER	Spacecraft event time	DN
003	et	FLOAT64	ephemeris time	DN

004	moon_id	INTEGER	Navigation and Ancillary Information Facility moon ID	DN
005	moon_prob	FLOAT64	Moon-bound probability	DN
X01	q_moon_SUFFIX	FLOAT64	Perihelion	SR
X02	e_moon_SUFFIX	FLOAT64	Eccentricity	DN
X03	a_moon_SUFFIX	FLOAT64	Semi-major axis	SR
X04	i_moon_SUFFIX	FLOAT64	Inclination	deg
X05	lnode_moon_SUFFIX	FLOAT64	Long. of asc. node	deg
X06	argp_moon_SUFFIX	FLOAT64	Argument of periapsis	deg
X07	mzero_moon_SUFFIX	FLOAT64	True anomaly at epoch	deg
X08	tzero_moon_SUFFIX	FLOAT64	Corresponding epoch	s
006	conics_moon_cov	BLOB	Covariance matrix of the orbital elements	misc

Table: saturn_ini

Table C.14 shows the content of table *saturn_ini*. It contains dust particles on an escape orbit with respect to Saturn. Consequently, also exogenic particles are included, with respect to the Saturnian system.

Table C.14.: Entries of the orbit database table *saturn_ini*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity. Orbital elements are given with respect to SCNRC.

CN	Column Name	Data Type	Description	Unit
001	SCLK	INTEGER	Spacecraft clock	DN
002	SCET	INTEGER	Spacecraft event time	DN
003	et	FLOAT64	ephemeris time	DN
004	saturn_inbound_prob	FLOAT64	Saturn-inbound probability	DN
005	saturn_outbound_prob	FLOAT64	Saturn-outbound probability	DN
X01	q_saturn_ini_SUFFIX	FLOAT64	Perihelion	SR
X02	e_saturn_ini_SUFFIX	FLOAT64	Eccentricity	DN
X03	a_saturn_ini_SUFFIX	FLOAT64	Semi-major axis	SR
X04	i_saturn_ini_SUFFIX	FLOAT64	Inclination	deg
X05	lnode_saturn_ini_SUFFIX	FLOAT64	Long. of asc. node	deg
X06	argp_saturn_ini_SUFFIX	FLOAT64	Argument of periapsis	deg
X07	mzero_saturn_ini_SUFFIX	FLOAT64	True anomaly at epoch	deg
X08	tzero_saturn_ini_SUFFIX	FLOAT64	Corresponding epoch	s

006	conics_saturn_ini_cov	BLOB	Covariance matrix of the orbital elements	misc

Table: moon_ini

Table C.15 shows the content of table *moon_ini*. It contains dust particles on an escape orbit with respect to a moon. Consequently, exogenic and Saturn-bound particles are included, with respect to the corresponding moon system.

Table C.15.: Entries of the orbit database table *moon_ini*. The first column is a column number, the second is the column name in the database, the third column gives a brief description and the last one shows the physical quantity. Orbital elements are given with respect to MCNRC.

CN	Column Name	Data Type	Description	Unit
001	SCLK	INTEGER	Spacecraft clock	DN
002	SCET	INTEGER	Spacecraft event time	DN
003	et	FLOAT64	ephemeris time	DN
004	moon_id	INTEGER	Navigation and Ancillary Information Facility moon ID	DN
005	moon_ini_prob	FLOAT64	Moon-ini probability	DN
006	moon_inbound_prob	FLOAT64	Moon-inbound probability	DN
007	moon_outbound_prob	FLOAT64	Moon-outbound probability	DN
008	moon_source_prob	FLOAT64	Moon-source probability	DN
X01	q_moon_ini_SUFFIX	FLOAT64	Perihelion	SR
X02	e_moon_ini_SUFFIX	FLOAT64	Eccentricity	DN
X03	a_moon_ini_SUFFIX	FLOAT64	Semi-major axis	SR
X04	i_moon_ini_SUFFIX	FLOAT64	Inclination	deg
X05	lnode_moon_ini_SUFFIX	FLOAT64	Long. of asc. node	deg
X06	argp_moon_ini_SUFFIX	FLOAT64	Argument of periapsis	deg
X07	mzero_moon_ini_SUFFIX	FLOAT64	True anomaly at epoch	deg
X08	tzero_moon_ini_SUFFIX	FLOAT64	Corresponding epoch	s
009	conics_moon_ini_cov	BLOB	Covariance matrix of the orbital elements	misc
X09	source_moon_ini_lon_SUFFIX	FLOAT64	Moon-surface source coordinates, longitude	deg
X10	source_moon_ini_lat_SUFFIX	FLOAT64	Moon-surface source coordinates, longitude	deg

X11	source_moon_ini_vel_SUFFIX	FLOAT64	Moon-surface source coordinates, longitude	deg
010	source_moon_ini_cov	FLOAT64	Covariance matrix of moon-surface source coordinates	deg

C.5. Neural Network (NN) velocity calibration

C.5.1. Neural Network Regression Models - Fitting

Table C.16.: Neural Network regression results for the Chemical Analyzer Target QC data. The first and second column list the model and dropout rate. However the last row shows the results for the Gaussian Noise layers and gives the σ noise. Column 3 and 4 show the batchsize and epoch for the best regression model. The according score and standard deviation of the mean score is shown in the last two columns. Both parameters are given in the scaled domain.

Model	Dropout	Batchsize	Epochs	μ Score (scaled)	σ Score (scaled)
Simple	—	15	8	0.286	0.032
Default	0.1	20	16	0.273	0.022
Deep	0.0	15	64	0.268	0.023
Adaptive (Dropout)	0.0	25	8	0.269	0.042
Adaptive (GDropout)	0.1	30	64	0.261	0.033
	σ Noise				
Adaptive (GNoise)	0.0	30	128	0.273	0.044

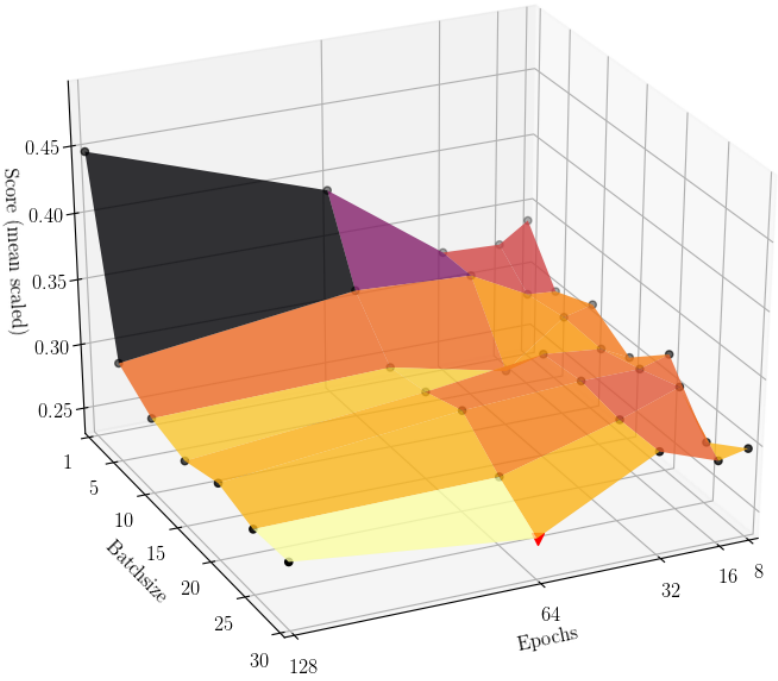


Figure C.5.: Score domain for the Chemical Analyzer Target QC regression model (*Adaptive* model with a 0.1 Gaussian dropout rate). The scaled mean score is plotted vs. the used batchsize and number of epochs. For better readability, the dots that represent the scores and connected with height depending color planes, where brighter colors indicate a better score and vis versa. The best solution is indicated with a red triangle.

Table C.17.: Neural Network regression results for the Chemical Analyzer Target QI data. The first and second column list the model and dropout rate. However the last row shows the results for the Gaussian Noise layers and gives the σ noise. Column 3 and 4 show the batchsize and epoch for the best regression model. The according score and standard deviation of the mean score is shown in the last two columns. Both parameters are given in the scaled domain.

Model	Dropout	Batchsize	Epochs	μ Score (scaled)	σ Score (scaled)
Simple	—	20	32	0.38	0.12
Default	0.1	25	32	0.349	0.047
Deep	0.0	30	16	0.348	0.047
Adaptive (Dropout)	0.1	30	16	0.339	0.046
Adaptive (GDropout)	0.0	20	16	0.331	0.044
	σ Noise				
Adaptive (GNoise)	0.0	30	32	0.34	0.036

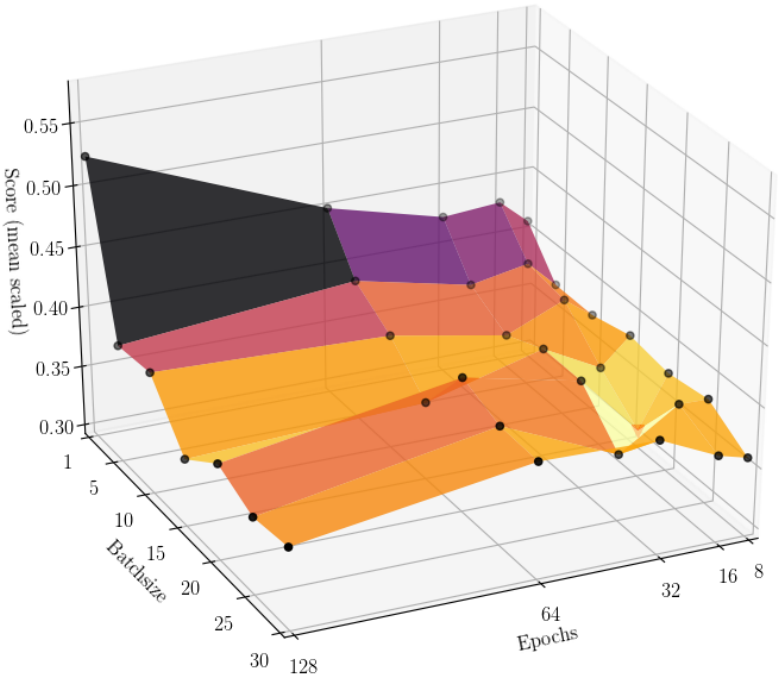


Figure C.6.: Score domain for the Chemical Analyzer Target QI regression model (*Adaptive* model with a 0.0 Gaussian dropout rate). The scaled mean score is plotted vs. the used batchsize and number of epochs. For better readability, the dots that represent the scores and connected with height depending color planes, where brighter colors indicate a better score and vis versa. The best solution is indicated with a red triangle.

Table C.18.: Neural Network regression results for the Impact Ionisation Detector QT data. The first and second column list the model and dropout rate. However the last row shows the results for the Gaussian Noise layers and gives the σ noise. Column 3 and 4 show the batchsize and epoch for the best regression model. The according score and standard deviation of the mean score is shown in the last two columns. Both parameters are given in the scaled domain.

Model	Dropout	Batchsize	Epochs	μ Score (scaled)	σ Score (scaled)
Simple	—	25	8	0.52	0.13
Default	0.0	30	8	0.43	0.11
Deep	0.0	30	64	0.46	0.12
Adaptive (Dropout)	0.0	25	16	0.43	0.13
Adaptive (GDropout)	0.0	25	8	0.41	0.086
	σ Noise				
Adaptive (GNoise)	0.1	30	16	0.44	0.12

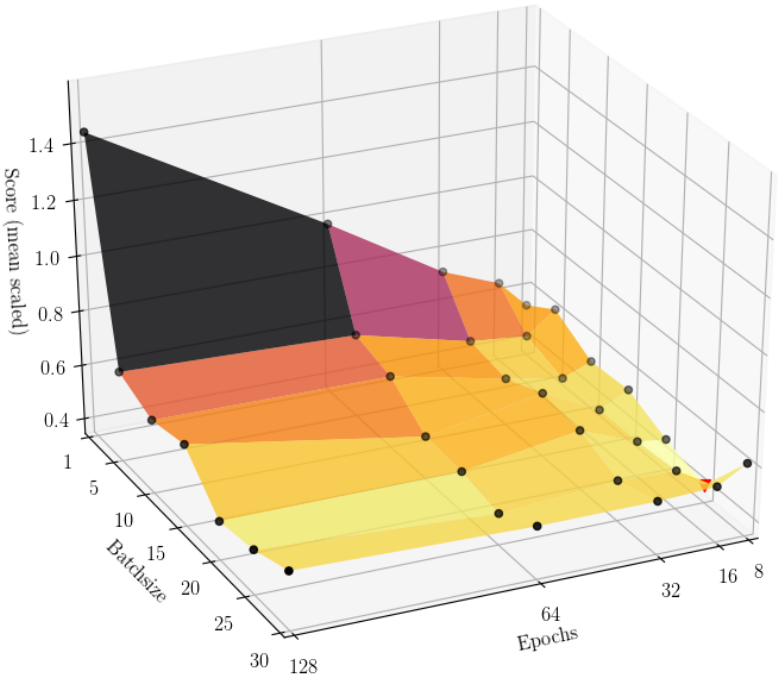


Figure C.7.: Score domain for the Chemical Analyzer Target QI regression model (*Adaptive* model with a 0.0 Gaussian dropout rate). The scaled mean score is plotted vs. the used batchsize and number of epochs. For better readability, the dots that represent the scores and connected with height depending color planes, where brighter colors indicate a better score and vis versa. The best solution is indicated with a red triangle.

Table C.19.: Neural Network regression results for Wall QI data. The first and second column list the model and dropout rate. However the last row shows the results for the Gaussian Noise layers and gives the σ noise. Column 3 and 4 show the batchsize and epoch for the best regression model. The according score and standard deviation of the mean score is shown in the last two columns. Both parameters are given in the scaled domain.

Model	Dropout	Batchsize	Epochs	μ Score (scaled)	σ Score (scaled)
Simple	—	15	8	0.56	0.24
Default	0.3	25	16	0.5	0.24
Deep	0.3	5	32	0.47	0.19
Adaptive (Dropout)	0.3	15	64	0.45	0.17
Adaptive (GDropout)	0.5	15	16	0.49	0.26
	σ Noise				
Adaptive (GNoise)	0.1	25	8	0.45	0.19

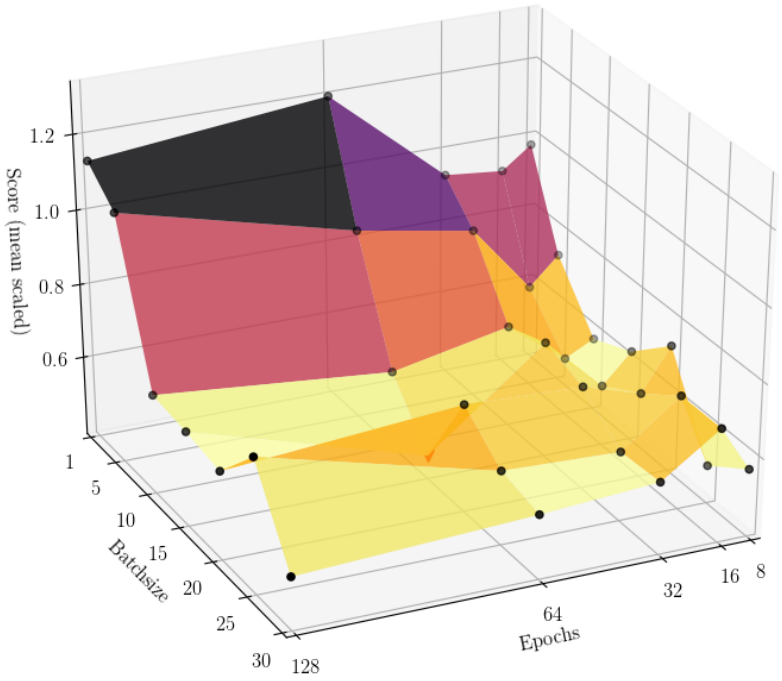


Figure C.8.: Score domain for the Wall QI regression model (*Adaptive* model with a 0.3 dropout rate). The scaled mean score is plotted vs. the used batchsize and number of epochs. For better readability, the dots that represent the scores and connected with height depending color planes, where brighter colors indicate a better score and vis versa. The best solution is indicated with a red triangle.

C.5.2. Final Neural Network Regression Model (NNRM) models

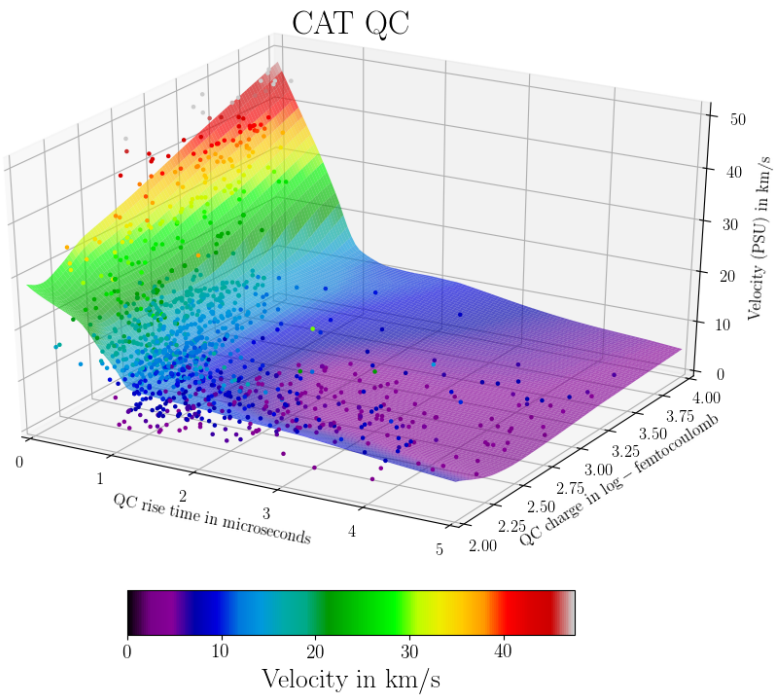


Figure C.9.: Chemical Analyzer Target QC calibration data with the computed neural network regression model (plane). The velocity in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

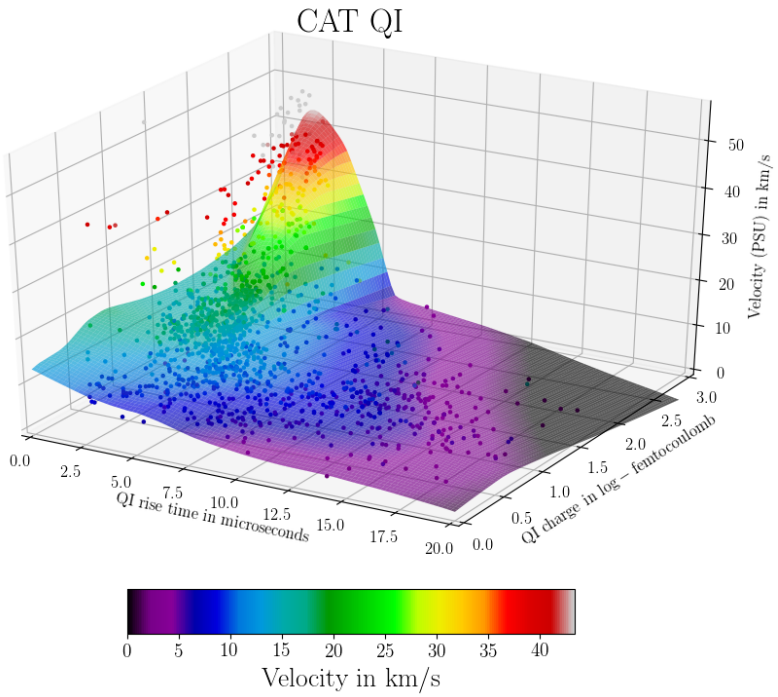


Figure C.10.: Chemical Analyzer Target QI calibration data with the computed neural network regression model (plane). The velocity in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

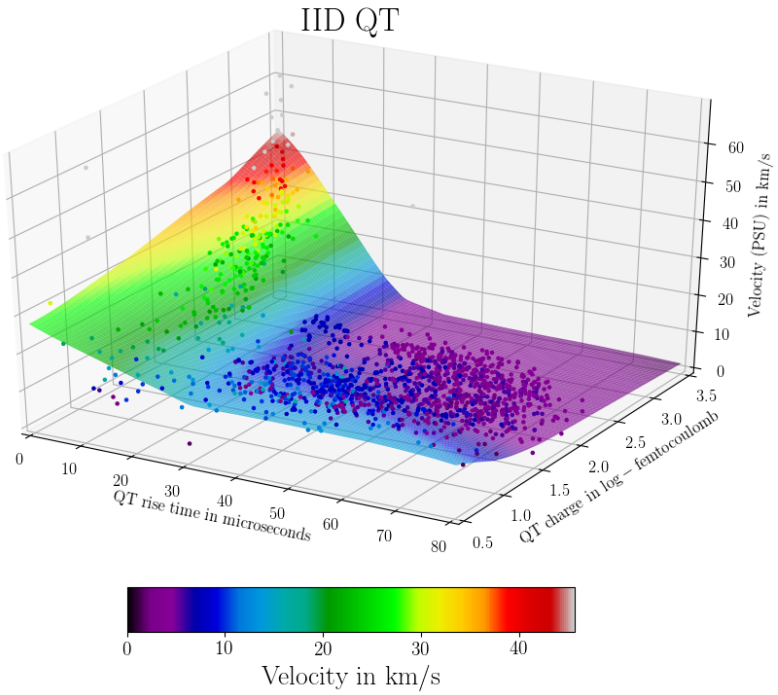


Figure C.11.: Impact Ionisation Detector QT calibration data with the computed neural network regression model (plane). The velocity in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

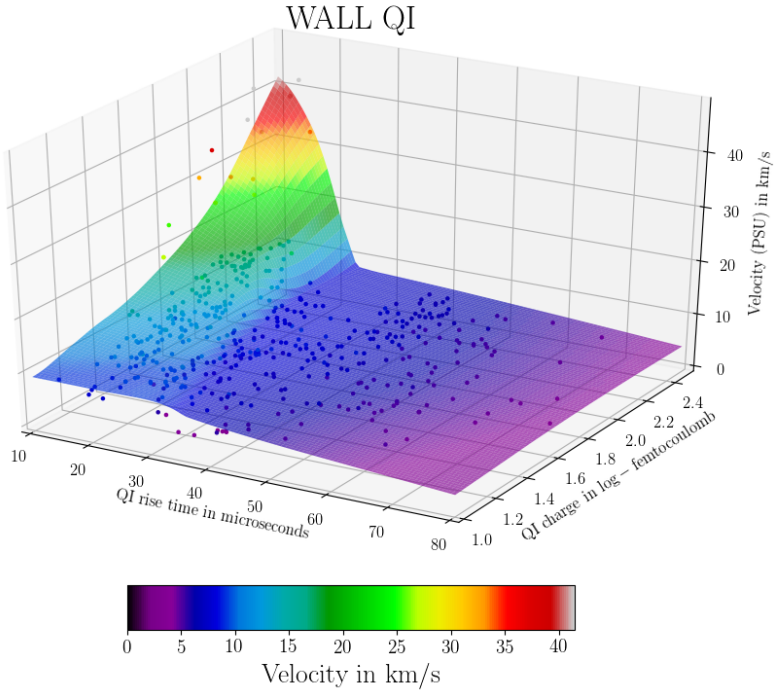


Figure C.12.: Wall QI calibration data with the computed neural network regression model (plane). The velocity in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

C.5.3. Error Determination

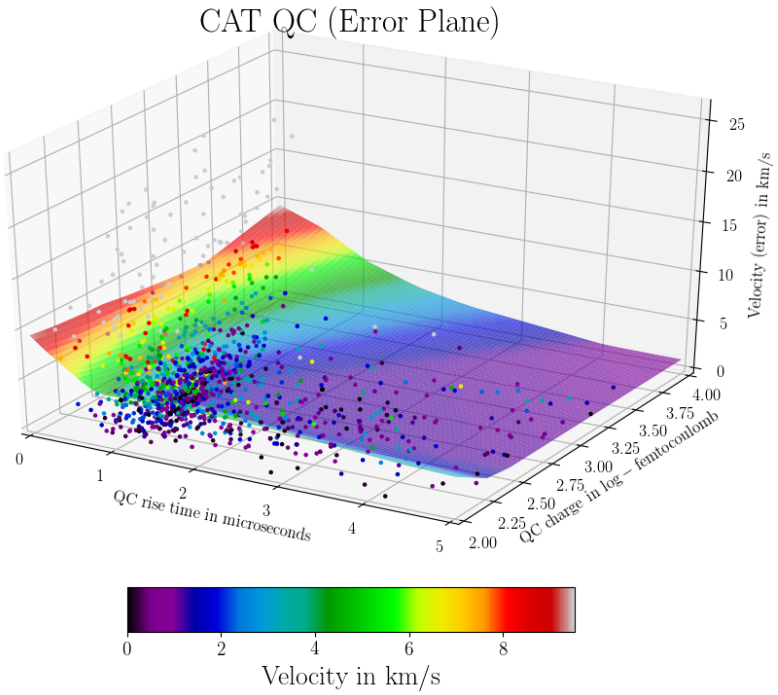


Figure C.13.: Chemical Analyzer Target QC deviation data with the computed error model (plane). The velocity deviation in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

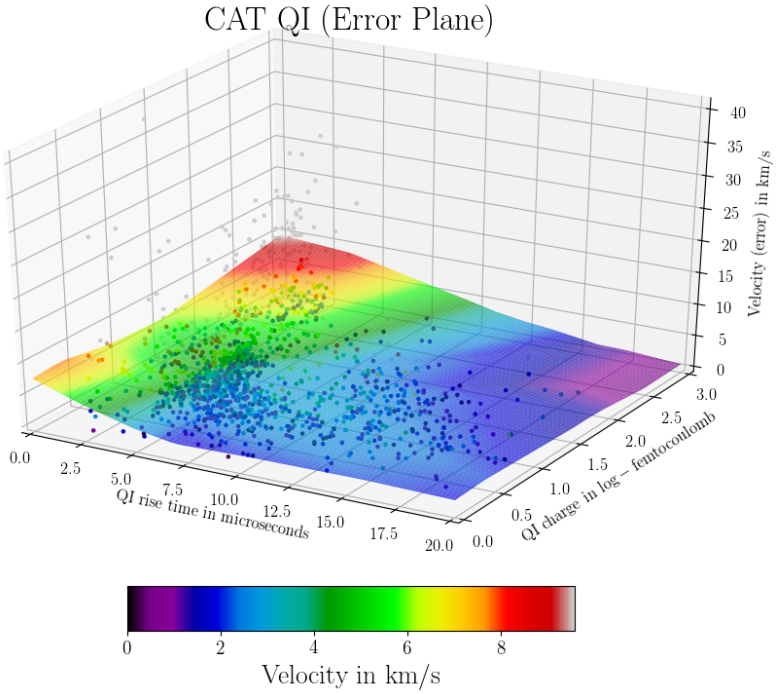


Figure C.14.: Chemical Analyzer Target QI deviation data with the computed error model (plane). The velocity deviation in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

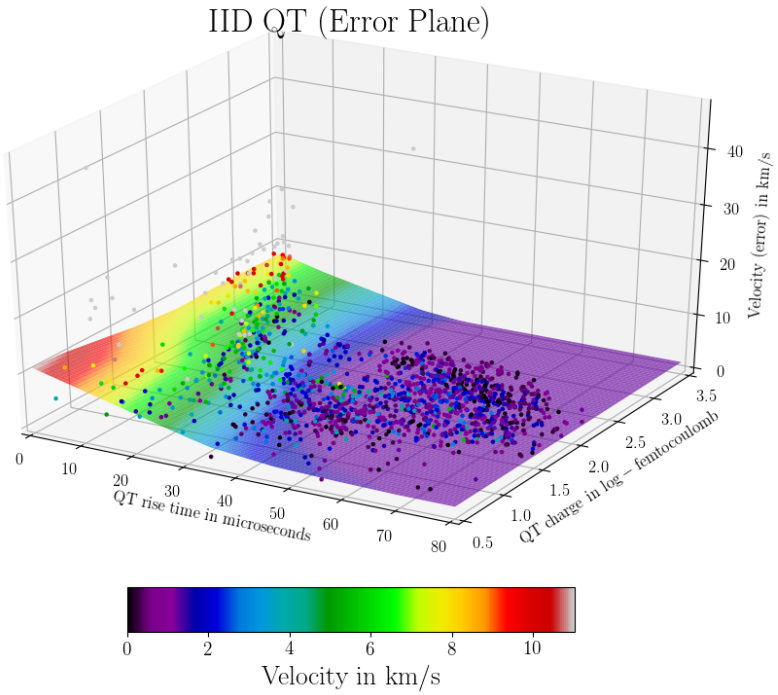


Figure C.15.: Impact Ionisation Detector QT deviation data with the computed error model (plane). The velocity deviation in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

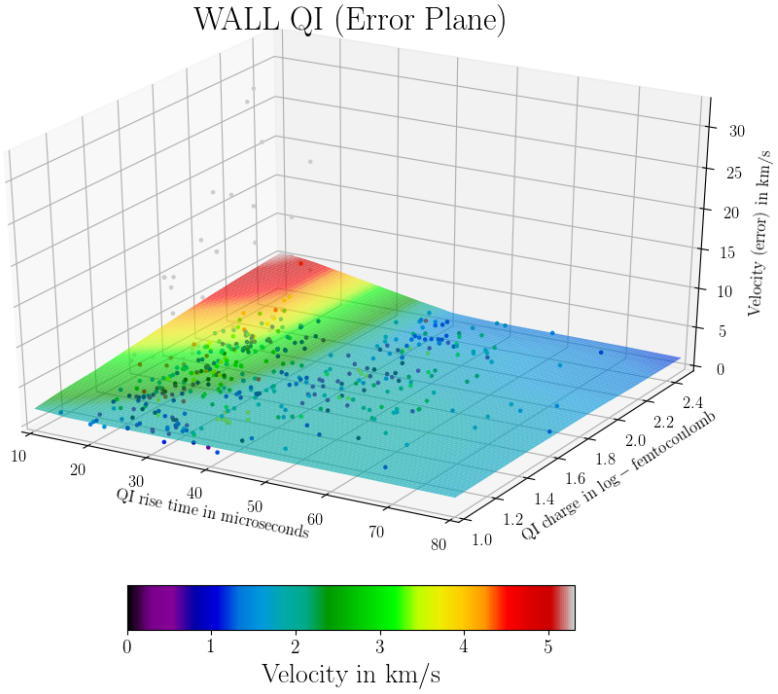
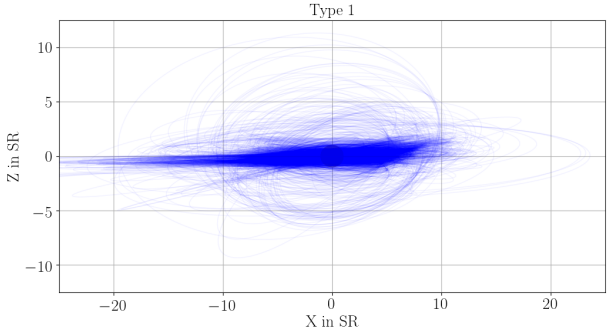


Figure C.16.: WALL QI deviation data with the computed error model (plane). The velocity deviation in km/s is plotted vs. the channel's rise time in μs and charge in $\log_{10}$ fC, respectively. For better readability, the velocity is also color-coded.

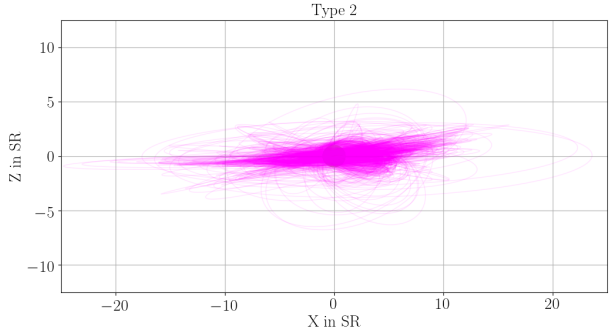
D. Machine learning based spectra classification

D.1. A-priori labeled orbit data

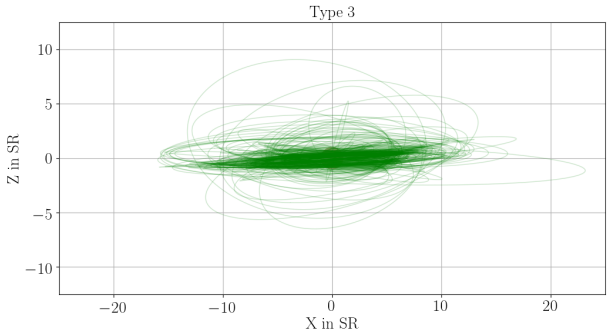
This section shows the orbits of the pre-classified dust detections (Noelle L. & Postberg F. (priv. comm.)). The plots are edge-on views on the Saturnian system for the different types as described in figure 5.77



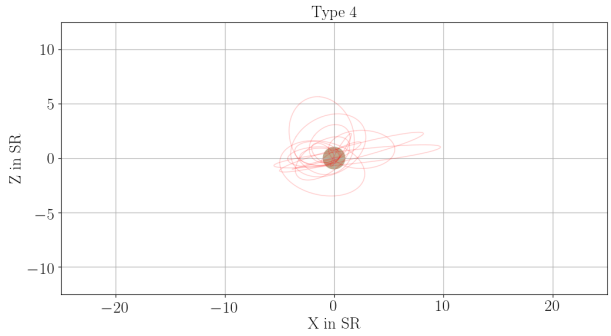
(a) Orbits of type 1 Saturn bound particles



(b) Orbits of type 2 Saturn bound particles

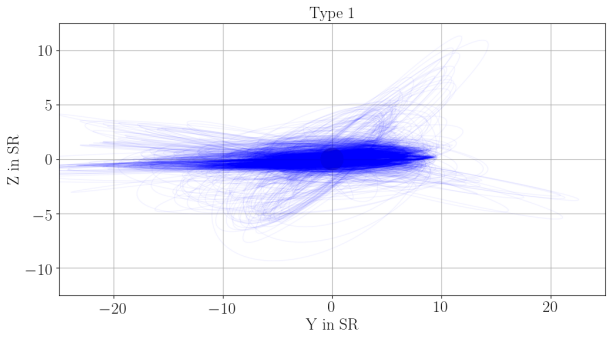


(c) Orbits of type 3 Saturn bound particles

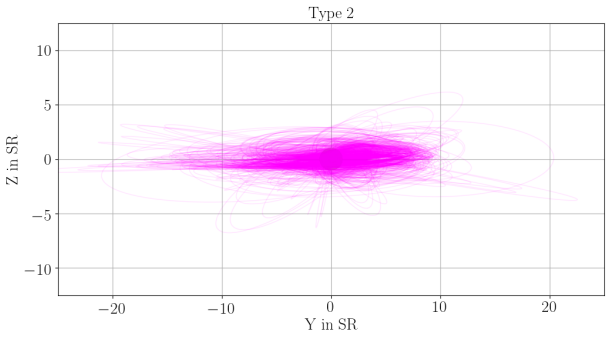


(d) Orbits of type 4 Saturn bound particles

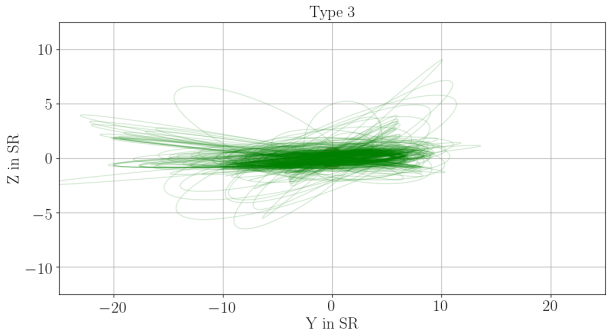
Figure D.1.: Orbits of the determined Saturn bound particles. Each subplot shows the orbits for each classified type. The coordinate system is SCNRC with Saturn and its ring in the center (edgeview: X-Z). The inner and outer dashed circles represent the orbits of Enceladus and Titan, respectively.



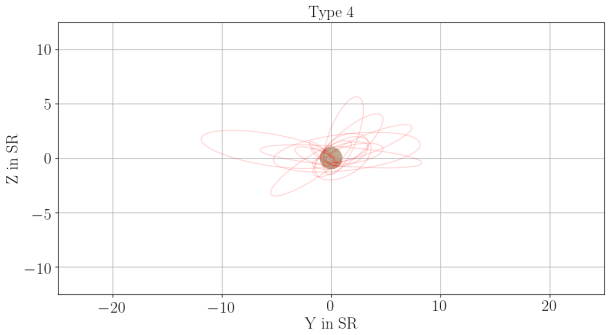
(a) Orbits of type 1 Saturn bound particles



(b) Orbits of type 2 Saturn bound particles



(c) Orbits of type 3 Saturn bound particles



(d) Orbits of type 4 Saturn bound particles

Figure D.2.: Orbits of the determined Saturn bound particles. Each subplot shows the orbits for each classified type. The coordinate system is SCNRC with Saturn and its ring in the center (edgeview: Y-Z). The inner and outer dashed circles represent the orbits of Enceladus and Titan, respectively.