Universität Regensburg

Estimation of Dose Distribution for Lu-177 Therapies in Nuclear Medicine

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Fakultät für Sprach-, Literatur und Kulturwissenschaft
der Universität Regensburg vorgelegt von

Dr. rer. nat. Theresa Ida Götz
aus Weiden in der Oberpfalz

2020, Fakultät für Sprach-, Literatur und Kulturwissenschaft

Die Arbeit entstand in gemeinsamer Betreuung durch die Fakultät für Physik der
Universität Regensburg und der Fakultät für Medizin der
Friedrich-Alexander-Universität Erlangen-Nürnberg

Regensburg, 2020

Erstgutachter (Betreuer): Prof. Dr.-Ing. Bernd Ludwig
Zweitgutachter: Prof. Dr. Elmar Wolfgang Lang
Drittgutachter: PD Dr. David Elsweiler

Abstract

In nuclear medicine, two frequent applications of ^{177}Lu therapy exist: DOTATOC therapy for patients with a neuroendocrine tumor and PSMA thearpy for prostate cancer. During the therapy a pharmaceutical is injected intravenously, which attaches to tumor cells due to its molecular composition. Since the pharmaceutical contains a radioactive ^{177}Lu isotope, tumor cells are destroyed through irradiation. Afterwards the substance is excreted via the kidneys. Since the latter are very sensitive to high energy radiation, it is necessary to compute exactly how much radioactivity can be administered to the patient without endangering healthy organs. This calculation is called dosimetry and currently is made according to the state of the art MIRD method. At the beginning of this work, an error assessment of the established method is presented, which has determined an overall error of 25% in the renal dose value. The presented study improves and personalizes the MIRD method in several respects and reduces individual error estimates considerably.

In order to be able to estimate of the amount of activity, first a test dose is injected to the patient. Subsequently, after 4h, 24h, 48h and 72h SPECT images are taken. From these images the activity at each voxel can be obtained a specified time points, i. e. the physical decline and physiological metabolization of the pharmaceutical can be followed in time. To calculate the amount of decay in each voxel from the four SPECT registrations, a time activity curve must be integrated. In this work, a statistical method was developed to estimate the time dependent activity and then integrate a voxel-by-voxel time-activity curve. This procedure results in a decay map for all available 26 patients (13 PSMA/13 DOTATOC).

After the decay map has been estimated, a full Monte Carlo simulation has been carried out on the basis of these decay maps to determine a related dose distribution. The simulation results are taken as reference ("Gold Standard") and compared with methods for an approximate but faster estimation of the dose distribution. Recently, a convolution with Dose Voxel Kernels (DVK) has been established as a standard dose estimation method (Soft Tissue Scaling STS). Thereby a radioactive Lutetium isotope is placed in a cube consisting of soft tissue. Then radiation interactions are simulated for a number of 10^{10} decays. The resulting Dose Voxel Kernel is then convolved with the estimated decay map. The result is

a dose distribution, which, however, does not take into account any tissue density differences. To take tissue inhomogeneities into account, three methods are described in the literature, namely Center Scaling (CS), Density Scaling (DS), and Percentage Scaling (PS). However, their application did not improve the results of the STS method as is demonstrated in this study. Consequently, a neural network was trained finally to estimate DVKs adapted to the respective individual tissue density distribution. During the convolution process, it uses for each voxel an adapted DVK that was deduced from the corresponding tissue density kernel. This method outperformed the MIRD method, which resulted in an uncertainty of the renal dose between $-42.37 - 10.22\%$ an achieve a reduction in the uncertainty to a range between $-26.00\% - 7.93\%$. These dose deviations were calculated for 26 patients and relate to the mean renal dose compared with the respective result of the Monte Carlo simulation. In order to improve the estimates of dose distribution even further, a 3D 2D neural network was trained in the second part of the work. This network predicts the dose distribution of an entire patient. In combination with an Empirical Mode Decomposition, this method achieved deviations of only $-12.21\% - 2.13\%$. The mean deviation of the dose estimates is in the range of the statistical error of the Monte Carlo simulation.

In the third part of the work, a neural network was used to automatically segment the kidney, spleen and tumors. Compared to an established segmentation algorithm, the method developed in this work can segment tumors because it uses not only the CT image as input, but also the SPECT image.

Zusammenfassung

In der Nuklearmedizin gibt es zwei Anwendungen einer ^{177}Lu-Therapie, zum einen die DOTATOC-Therapie für Patienten mit einem neuroendokrinen Tumor und zum anderen die PSMA-Thearpie für Prostatakrebs. Bei dieser Therapieform wird intravenös ein Stoff gespritzt, welcher sich aufgrund der molekularen Zusammensetzung an Tumorzellen anlagert. Da an die Moleküle ein radioaktives Lu-Isotope gekoppelt ist, werden so Tumorzellen durch die Bestrahlung getötet. Mit der Zeit wird der Stoff über die Nieren ausgeschieden. Da es sich bei den Nieren um ein strahlen-sensitives Organ handelt, muss vor der Injektion genau berechnet werden, wie viel Aktivität dem Patienten gespritzt werden kann, ohne die gesunden Organe zu gefährden. Diese Berechnung nennt man Dosimetrie und wird laut Stand der Technik mittels MIRD Verfahren durchgeführt. Zu Beginn dieser Arbeit wird eine Fehlerabschätzung der etablierten Methode vorgestellt, welche einen Gesamtfehler von 25% beim Nierendosiswert ermittelt hat.

Um eine Abschätzung der Aktivitätsmenge treffen zu können, wird dem Patienten zunächst eine Testdosis gespritzt. Anschließend werden nach 4h, 24h, 48h und 72h SPECT-Bilder aufgenommen. Aus den Bildern kann die Aktivität pro Voxel zu einen gewissen Zeitpunkt abgelesen werden. Um nun aus den vier Momentaufnahmen zu berechnen, wie viel Zerfälle pro Voxel stattgefunden haben, muss über eine Zeit-Aktivitätskurve integriert werden. In dieser Arbeit wurde ein statistisches Verfahren entwickelt, um eine voxelweise Zeit-Aktivitätskurve abzuschätzen und anschließend zu integrieren. Resultat dieses Verfahrens ist eine Zerfallskarte für jeden der 26 vorhandenen Patienten (13 PSMA/13 DOTATOC).

Nachdem die Zerfallskarte berechnet wurde, kann anhand dieser eine volle Monte Carlo Simulation erfolgen, welche eine Dosisverteilung ermittelt. Diese wird als Gold Standard angenommen und mit den Verfahren zur schnelleren Abschätzung der Verteilung verglichen. Als Standard Methode (Soft Tissue Scaling STS) wurde eine Faltung mit einem Dose Voxel Kernel (DVK) publik. Dabei wird in einem Würfel bestehend aus Weichteilgewebe ein Lutetium-Isotope platziert, anschließend werden 10^{10} Zerfälle simuliert. Der resultierende Dose Voxel Kernel wird dann mit der Zerfallskarte gefaltet. Resultat ist eine Dosisverteilung, welche jedoch keine Dichteunterschiede berücksichtigt. Um Letztere zu berück-

sichtigen, existieren in der Literatur drei Methoden, nämlich Center Scaling (CS), Density Scaling (DS) und Percentage Scaling (PS). Diese erzielten jedoch in dem in der Arbeit durchgeführten Vergleich keine Verbesserung zur STS Methode. Um auf die jeweilige Dichteverteilung angepasste DVKs abzuschätzen, wurde ein neuronales Netz trainiert. Dieses verwendet während des Faltungsprozess für jeden Voxel einen anderen DVK, der für diesen Dichtekernel vorhergesagt wurde. Durch diese Methode konnte statt $-42.37\% - 10.22\%$ Abweichung in der Nierendosis eine reduzierte Range von $-26.00\% - 7.93\%$ erzielt werden. Die Abweichungen sind ein Mittelwert über 26 Patienten und betreffen die mittlere Nierendosis im Vergleich mit der Monte Carlo Simulation. Um die Schätzungen der Dosisverteilung noch zu verbessern, wurde im zweiten Teil der Arbeit ein 3D-2D-Neural-Network trainiert. Dieses Netz sagt die Ganzkörper-Dosisverteilung eines gesamten Patienten vorher. In Kombination mit einer Empirical Mode Decomposition erzielte diese Methode Abweichungen von nur $-12.21\% - 2.13\%$. Diese mittlere Abweichung der Dosisschätzungen liegt im Bereich des statistischen Fehlers der Monte Carlo Simulation.

Im dritten Teil der Arbeit wurde ein Neuronales Netz verwendet um automatisiert die Niere, Milz und Tumore zu segmentieren. Im Vergleich zu einem etablierten Segmentierungsalgorithmus kann die in dieser Arbeit entwickelte Methode Tumore segmentieren, da sie nicht nur das CT Bild als Eingabe verwendet, sondern auch das SPECT Bild.

Contents

Abbreviations **9**

Introduction **12**

1 Nuclear Medical Physics **17**
1.1 ^{177}Lu radionuclide therapies 17
1.1.1 Neuroendocrine tumors (NETs) 17
1.1.2 Prostate cancer (PC) 18
1.2 The Dosimetry Chain for Radionuclide Therapy 19
1.2.1 Time-Activity-Curve (TAC) 22
1.2.2 Energy dose rate and energy dose 23
1.2.3 Radiation energy transport and dose kernels 25
1.2.4 MIRD-method 26
1.2.5 Voxelwise dosimetry 28
1.3 Uncertainties in dose estimation 30

2 Monte Carlo Simulation (MC) **31**
2.1 Radionuclide 33
2.2 Standard Error for Dose Estimation 34

3 Particle filter (PF) **41**
3.1 State Estimation 44
3.1.1 The model equations 44
3.1.2 The simplifying assumptions 45
3.1.3 The problems to be solved 45
3.2 The Kalman Filter 46
3.3 Bayesian Filters 47
3.4 Particle Filter 48
3.4.1 Sampling on a dynamic stochastic grid 48
3.4.2 Sequential importance sampling 50
3.4.3 Resampling Schemes 55

4 Bi-dimensional EMD **59**
4.1 Canonical Bi-dimensional EMD 60
4.2 Noise-assisted BEMD 62
4.3 Green's function-based BEEMD 63
4.3.1 Extraction of local extrema 64
4.4 Green's function for estimating envelopes 64
4.5 Dimension reduction by PCA 68

5 Neural networks (NNs) **69**
5.1 Gradient descent optimization in NNs 69
5.1.1 General 69
5.1.2 Gradient descent 70
5.1.3 RMSProp 71
5.2 Deep Learning 72
5.2.1 2D Convolutions 72
5.2.2 ReLUs in DNNs 76
5.3 Convolutional neural networks (CNNs) 77
5.3.1 Architecture of CNNs 77
5.4 U-Nets 80
5.4.1 Network Architecture 82
5.4.2 Training 83

6 Dataset **87**
6.1 Patient collective 87
6.2 Image Acquisition 88
6.3 Common Preprocessing 90

7 Error of the standard MIRD method **93**
7.1 Segmentation of an organ on CT 93
7.2 Registration of 4 SPECT-images on CT 93
7.3 Amount of activity within organ 94
7.4 Integration of time-activity-curve (TAC) 94
7.5 Scale standard phantom and multiply with S-Value 95
7.6 Results 95
7.7 Discussion 96

8 Time-integrated activity map (TIA) **99**
8.1 Model equations 99
8.2 Other Methods 101
8.3 Results 102
8.3.1 Parameter evaluation 102

8.3.2 Voxelwise TAC . . . 103
8.3.3 Whole patient . . . 104
8.4 Discussion . . . 105

9 Dose estimations 109
9.1 Ground truth: Monte Carlo simulations (MC) . . . 109
9.1.1 Dose Voxel Kernel (DVK) . . . 111
9.1.2 Statistical error of a whole body simulation . . . 113
9.1.3 Homogeneous patient . . . 113
9.2 Dose estimation with DVK via Scaling . . . 115
9.2.1 Soft Tissue Scaling (STS) . . . 116
9.2.2 Center Scaling (CS) . . . 116
9.2.3 Density Scaling (DS) . . . 116
9.2.4 Percentage scaling (PS) . . . 117
9.3 Dose estimation with DVK via NN (DVK-NN) . . . 118
9.3.1 Network architecture . . . 118
9.3.2 Training data and evaluation . . . 120
9.4 Dose estimation of whole patients via NN . . . 123
9.4.1 Preprocession . . . 123
9.4.2 Net architecture . . . 124
9.5 Comparison of different methods . . . 128
9.5.1 Results . . . 128
9.5.2 Discussion . . . 130
9.6 Organ dose calculation via deep neural networks . . . 132
9.6.1 Data statistics . . . 132
9.6.2 Network architecture . . . 133
9.6.3 Results . . . 133

10 Medical Achievements 137
10.1 Statistical analysis . . . 137
10.2 PSMA . . . 137
10.3 DOTATOC . . . 140
10.4 Limitations . . . 143

Conclusion 143

Publications based on this study 147

Literature 171

Acknowledgements 173

Abbreviations

SPECT/CT *Single-Photon Emission Computed Tomography / CT*
PET/CT *Positron-Emission Tomography / CT*
SPECT *Single-Photon Emission Computed Tomography*
PET *Positron-Emission Tomography*
CT *Computed Tomograpy*
MIRD *Medical Internal Radiation Dose Committee*
DVK *Dose Voxel Kernel*
DPK *Dose Point Kernel*
MC *Monte Carlo Simulation*
NET *Neuroendocrine tumor*
ADT *Androgen Deprivation Therapy*
PRRT *Peptide Radionuclide Receptor Therapy*
mCSPC *metastatic castration-sensitive* Prostate Cancer *(PC)*
mCRPC *metastatic castration-resistant PC*
SSTR *Somatostatin Receptor*
TTD *Total Tumor Dose*
PC *Prostate Cancer*
PSMA *Prostate Specific Membrane Antigen*
TAC *Time Activity Curve*
EM *Expectation Maximization*
TIA *Time Integrated Activity*
VOI *Volume Of Interest*
OLINDA *Organ Level Internal Dose Assessment with Exponential Modeling*
SE *Standard Error*
SD *Standard Deviation*
SS *Sample Size*
PF *Particle Filter*
SNR *Signal to Noise Ratio*
DNN *Deep Neural Networks*
EMD *Empirical Mode Decomposition*
EEMD *Ensemble Empirical Mode Decomposition*

MEEMD *Multi-dimensional Ensemble Empirical Mode Decomposition*
BEMD *Bi-Dimensional EMD*
CNN *Convolutional Neural Network*
IMF *Intrinsic Mode Function*
BIMF *Bi-dimensional Intrinsic Mode Function*
BEEMD *Bi-dimensional* Ensemble Empirical Mode Decomposition *(EEMD)*
GiT-BEMD *Green's function in Tension* - Bi-dimensional EEMD *(BEEMD)*
BIM *Bi-dimensional Intrinsic Mode*
MLE *Maximum Likelihood Estimation*
ReLU *Leaky Rectified Linear Units*
MLP *Multi-Layer Perceptrons*
LCN *Local Contrast Normalization*
SGD *Stochastic Gradient Descent*
KL *Kullback-Leibler divergence*
CgA *Chromogranin A*
PSA *Prostate Specific Antigen*
AUC *Area Under the Curve*
SEM *State Evolution Model*
OM *Observation Model*
HU *Hounsfield Units*
STS *Soft Tissue Scaling*
CS *Center Scaling*
DS *Density Scaling*
PS *Percentage Scaling*
DVK-NN *DVK via Neural Networks*
NN *Neural Network*
DE-NN-EMD *Dose Estimation via Neural Network and EMD*
DE-NN *Dose Estimation via Neural Network*
LOOCV *Leave-One-Out Cross-Validation*
RLT *Radioligand Therapy*
FBP *Filtered Back-Projection*
SIR *Sampling Importance Resampling*
PCA *Principal Component Analysis*
CMSC *Comparable Minimal Scale Combination Principle*
GT *Ground truth*

Introduction

More and more therapies with radiopharmaceuticals, labeled with beta emitting isotopes, are established, making the need for a patient-specific dosimetry of fundamental importance [22, 157, 55]. With the aid of hybrid *Single-Photon Emission Computed Tomography / CT* (SPECT/CT) and *Positron-Emission Tomography / CT* (PET/CT) imaging techniques, it is possible to obtain tissue density information in conjunction with the distribution of radioactivity inside the human body. The spatial resolution of nuclear medicine imaging of such activity distributions is in the mm to cm range. Consequently, macroscopic non-uniformities in the activity distribution can be detected. In [141, 53] an overview of imaging-based patient-specific dosimetry methods is given.

First of all, a spatial distribution of radioactivity over time using the patient's own anatomy is needed to obtain as output the spatial distribution of absorbed energy dose. [142, 69] A well-known dosimetry method is the calculation of S-values according to the *Medical Internal Radiation Dose Committee* (MIRD) formalism, where a standard phantom is defined [121]. The aim of this work is to obtain a patient specific dosimetry by using *Dose Voxel Kernel* (DVK).

The energy dose produced in a specific volume can be calculated as the linear superposition of contributions from each voxel (volume element) in the spatial activity distribution treated as a radiation point source. The energy dose produced by a radiation point source of isotropic unit activity in a homogeneous medium of infinite extension is called a *Dose Point Kernel* (DPK) [119, 18]. A number of publications compare DPK data bases generated with different *Monte Carlo Simulation* (MC) softwares such as GATE, FULKA, MCNP4C, CGSnrc and GEANT4 [115, 22]. The generally considered tissues are bone, lung, soft tissue and water. All data bases contain DPKs for the radioactive isotopes Iodine ^{131}I and Yttrium ^{90}Y. Instead of calculating a continuous DPK, a discrete DVK, also called S-voxel kernel, can be determined [113, 114]. The voxel size can be chosen arbitrarily, because the kernel can be scaled, as shown in [113, 125, 42]. To obtain the distribution of absorbed energy dose based on the patient's anatomy, the decay map has to be convolved with the DVK.

The estimation of DVKs is hardly explored in literate, yet some studies are

reported. Dieudonne at al. [36] report a comparison of dose estimations obtained with either a full MC simulation or by convolution with a DVK. The latter was computed for homogeneous soft tissue only and results have been reported for two patient specific dose estimations. In contrast, Scarinci et al. [137] computed DVKs for several tissues, but did not provide any comparison to the standard MIRD method. Hence possible improvements cannot be quantified, limiting the value of this study. Rather than computing DVKs for many tissues, results obtained with soft tissue density kernels could be scaled according to the tissue density under study. Along these lines, Dieudonne at al. [37] proposed that the product of the energy dose per voxel times the mass density of that voxel should be an invariant of the system. This proposal leads to the following density correction:

$$D_\rho(\mathbf{r}) \cdot \rho(\mathbf{r}) = D_{ST}(\mathbf{r}) \cdot \rho_{ST} \tag{1}$$

where $D_{ST}(\mathbf{r}) = \partial E(\mathbf{r})/\rho_{ST}\partial V_{vox}$ is the dose for one voxel with volume V_{vox}, centered at $\mathbf{r} = (x, y, z)^T$, and calculated for soft tissue as the material inside the voxel, $\rho(\mathbf{r})$ is the voxel mass density, $D_\rho(\mathbf{r})$ is the density-corrected voxel dose value and $\rho_{ST} = 1040\ kg/m^3$ is the mass density of soft tissue. This density scaling of absorbed energy dose was demonstrated on three different clinical cases. In each case, three different dose estimation methods were employed: a full MC simulation of the DVK, a convolution of the TIA map with a soft tissue DVK and finally a convolution with a density-scaled DVK

In yet another recent study, the DVKs where first scaled according to the energy dose distribution calculated in different tissues, and then convolved with the activity distribution. The aim was to take into account interface regions, where the mass density changes abruptly. This method is evaluated on the Zubal phantom for three different dose distributions. [76].

Similarly, the aim of the work reported in [100] was to evaluate the application of tissue-specific dose kernels instead of water dose kernels to improve the accuracy of patient-specific dosimetry by taking tissue heterogeneities into consideration. Tissue-specific DPKs and DVKs for Yttrium-90 (^{90}Y), Lutetium-177 (^{177}Lu), and Phosphorus-32 (^{32}P) were calculated using the MC code GATE (version 7). The calculated DPKs for bone, lung, adipose, breast, heart, intestine, kidney, liver, and spleen are compared with those of water. The dose distribution in normal and tumorous tissues in lung, liver, and bone of a Zubal phantom was calculated using tissue-specific DVKs instead of those of water in conventional methods. For a tumor defined in a heterogeneous region in the Zubal phantom, the absorbed dose was calculated using a proposed algorithm, taking tissue heterogeneity into account. The algorithm was validated against full MCs and indicated that the largest differences between water and other tissue-specific

DPKs occurred in bone, whereby the difference amounted to $^{90}Y : 12.2 \pm 0.6\ \%$, $^{32}P : 18.8 \pm 1.3\ \%$, and $^{177}Lu : 16.9 \pm 1.3\ \%$. The second largest discrepancy corresponded to the lung with $^{90}Y : 6.3 \pm 0.2\ \%$,$^{32}P : 8.9 \pm 0.4\ \%$, and $^{177}Lu : 7.7 \pm 0.3\ \%$. For ^{90}Y, the mean absorbed dose in tumorous and normal tissues was calculated using tissue-specific DVKs in lung, liver, and bone. The results were compared with doses calculated considering the Zubal phantom water equivalent and the relative differences were $4.50\%, 0.73\%$, and 12.23%, respectively. For the tumor in the heterogeneous region of the Zubal phantom that includes lung, liver, and bone, the relative difference between mean calculated dose in tumorous and normal tissues based on the proposed algorithm and the values obtained from full MC dosimetry was 5.18%. The authors concluded that their algorithm potentially enabled patient-specific dosimetry and improved estimates of the average absorbed dose of Yttrium-90 in a tumor located in lung, bone, and soft tissue interface by 6.98% compared with the conventional methods.

In [93] it was pointed out that the aim of personalized radiotherapy is clearly expressed in the EU directive 2013/59/EURATOM Article 56:

For all medical exposure of patients for radiotherapeutic purposes, exposures of target volumes shall be individually planned and their delivery appropriately verified taking into account that doses to non-target volumes and tissues shall be as low as reasonably achievable and consistent with the intended radiotherapeutic purpose of the exposure.

This statement formulates the goals onto which this thesis will also concentrate, namely first an individual planning of radiation energy delivery and second a verification of the energy dose absorbed in the target volume.

Chapter 1

Nuclear Medical Physics

1.1 ^{177}Lu radionuclide therapies

In the department of nuclear medicine of the University Hospital Erlangen, two radionuclide therapies are routinely performed with ^{177}Lu - labeled molecules. These therapies provide a treatment either for *Neuroendocrine tumor* (NET)s or for PC. In this work, the dose values for both diseases were evaluated.

1.1.1 Neuroendocrine tumors (NETs)

NETs are defined as epithel neoplasms with predominant neuroendocrine differentiation that can be observed arising from neuroendocrine cells throughout the body [78]. Data from the Surveillance, Epidemiology, and End Results database suggest that NETs are more prevalent than previously reported with 51% of NETs arising from the gastrointestinal tract, 27% from the lungs and 6% from the pancreas [169, 170]. NETs of the midgut commonly metastasize to the liver, the mesentery and the peritoneum. Clinically, they are regarded as functional if they are associated with symtoms of hormonal hypersecretion, the so called carcinoid syndrome, or non-functional if they are not associated with hormonal hypersecretion [153]. First-line systemic therapy is primarily based on somatostatin analogs which significantly lengthen time to tumor progression and improve control of hormonal secretion [127, 25]. Besides everolimus, a potent inhibitor of mammalian target of rapamycin, for the treatment of non functional neuroendocrine tumors, so far there have been no standard second-line systemic treatment options [170, 83]. However the recent food and drugs administration approvel of ^{177}Lu-DOTATATE for the treatment of *Somatostatin Receptor* (SSTR) positive gastroenteropancreatic tumors, based on the results from the phase 3 Neuroendocrine Tumors Therapy trial opens new perspectives for the treatment of NETs

[153]. ^{177}Lu emits beta-particels with a maximum energy of $149\ keV$, a maximum particle range of $2\ mm$ and has a physical half-life of $6.7\ d$. Besides beta-particles it also emits gamma photons of $208\ keV$, which can be directly used for uptake quantification by serial scintigraphy and *Single-Photon Emission Computed Tomography* (SPECT) [47]. Most of the clinical protocols rely on empirical criteria for choosing the administered activity and the number of cycles [20]. Special emphasis has to be placed on the absorbed doses for kidney and bone marrow, since they are considered as the dose limiting organs in ^{177}Lu- *Peptide Radionuclide Receptor Therapy* (PRRT) [133, 162]. Due to the large intrapatient and intralesion variability of tumor uptake in PRRT of NETs [30] it is of utmost importance to improve individualized therapy planning. Therefore methods for accurate dosimetry of tumorous- and non-tumorous tissue and determination of predictive factors that are associated with high uptake of ^{177}Lu-DOTATATE in NETs are needed. As yet only a few previously conducted studies reported the use of the MIRD scheme and the unit density sphere model from Olinda for calculation of tumor-absorbed doses in ^{177}Lu-DOTATATE therapy [30, 67]. The aims of the present study were to evaluate the use of three-dimensional tumor dosimetry based on particle filtering methods and MC to determine the *Total Tumor Dose* (TTD) and to find predictive factors that are associated with a high TTD in patients with SSTR-positive NETs that underwent ^{177}Lu-DOTATATE therapy.

1.1.2 Prostate cancer (PC)

Androgen Deprivation Therapy (ADT) is the mainstay of therapy in patients with locally advanced PC, biochemically recurrent disease after failure of local treatments and in patients with metastatic PC [74, 64]. Although initially being highly effective in most men with *metastatic castration-sensitive PC* (mCSPC), disease progression to mCSPC occurs in the majority of men within 2-3 years [56] and is associated with a poor prognosis and short overall survival [146]. Docetaxel and cabazitaxel are the only United States Food and Drug Administration approved chemotherapies for the treatment of *metastatic castration-resistant PC* (mCRPC) patients providing increased progression-free survival and also overall survival, especially in combination with ADT [70, 154]. Recently radionuclide therapy has gained increasing importance with the approval of the alpha emitting particle ^{223}Ra-dichloride for the treatment of mCSPC patients with symptomatic bone metastases, but without known visceral metastases demonstrating overall survival and prolonging the time to the first symptomatic skeletal event [116]. However about one third of patients suffering from advanced prostate cancer present with lymph node or visceral metastases which are unresponsive to bone-seeking radiopharmaceuticals [122]. Radioligand therapy using ^{177}Lu-labeled *Prostate Specific Membrane Antigen* (PSMA) (prostate specific membrane antigen) has been

proven to be an effective therapeutic option with a favorable toxicity profile in this heavily pretreated patient population [168, 23, 124, 3]. Since the PSMA expression of PC cells is directly correlated to androgen independence, metastasis formation and progression of disease [97] the PSMA-targeting theranostic concept offers advantages for diagnosis and also therapy. To achieve best therapeutic results in radioligand therapy accurate dosimetry is crucial to determine the optimal treatment activity resulting in tumor irradiation with the maximum absorbed dose without causing toxicity to critical organs. Besides the high and specific uptake of PSMA ligands in PC cells, different normal organs (e.g. kidney, salivary glands, bone marrow) exhibit tracer accumulation [33]. Furthermore due to the intra patient and intra-lesion variability of tumor uptake in ^{177}Lu-PSMA therapy [112] individualization of therapy planning is of utmost importance. A particular advantage for therapeutic dosimetry is the mode of decay of ^{177}Lu since it emits beta particles with a maximum energy of $149\ keV$ providing tumor irradiation but also gamma photons of $208\ keV$ allowing uptake quantification by serial scintigraphy and SPECT [158]. For ^{177}Lu-PSMA therapy as yet only a few previously conducted studies reported the use of the MIRD scheme and the unit density sphere model from Olinda for calculation of absorbed dose in normal organs and tumor lesions [112, 33]. The aim of this study was to evaluate the use of a novel three-dimensional tumor dosimetry based on particle filtering methods and Monte Carlo simulations to determine the TTD (Total Tumor Dose) and to find predictive factors that are associated with a high TTD in patients suffering from mCRPC that underwent ^{177}Lu-PSMA-617 therapy.

1.2 The Dosimetry Chain for Radionuclide Therapy

The dosimetry process, dealing with quantitative imaging and image-based dosimetry, i. e. correcting for photon attenuation, photon scatter, γ-camera limitations and voxel-based calculation of the absorbed energy dose, encompasses three major steps:

- Image analysis, including image registration and segmentation as well as classification of normal tissue versus tumors.
- Pharmakokinetics modeling applied to data obtained from images acquired at consecutive time points.
- This allows to compute the total amount of absorbed energy dose. In addition, biologically effective dose and normal-tissue control probability need to be considered. The total amount of absorbed energy dose can be calculated employing one of three different methods [21]:

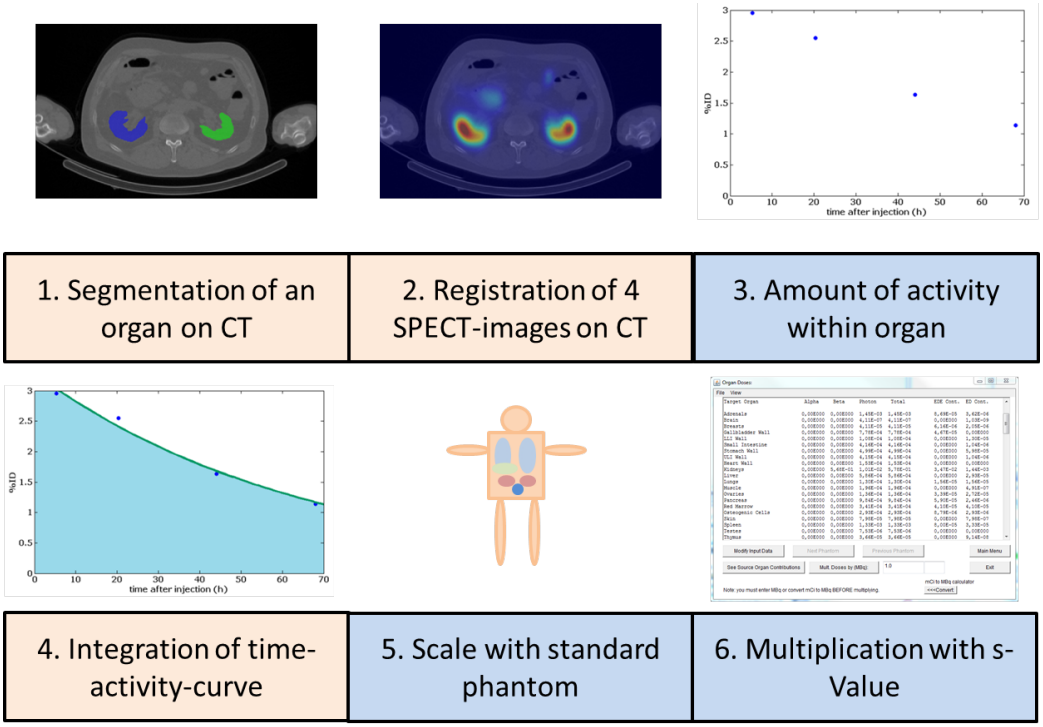


Figure 1.1: Steps of the dosimetry chain according to the MIRD protocol

- Multiplication of a global dose kernel, called S-value, with the cumulated activity according to MIRD
- Convolution of activity distributions with dose-voxel-kernels (DVK)
- Full Monte Carlo simulation (MC)

Any successful implementation has to rely on a close collaboration of medical physicists and technologists.

The order of the steps can vary, depending on the used dose calculation method. In figure 1.1 the steps for the MIRD-method are illustrated. For the voxel-wise methods (MC and DVK), three steps are identical but are applied in a different order as can be seen in figure 1.2.

The three dose estimation methods require the assessment of the 3D distribution of the radionuclide within the body. Over the last few years, efforts moved towards image processing methods to quantify a spatial and temporal activity distribution with good accuracy [48], [90], [27], [26]. With the advent of the latest PET/CT and SPECT/CT technologies, quantification of activity distributions can be performed with resolutions adequate for voxel dosimetry, typically of $5\ mm$ or even less, co-registering functional and anatomical images. The information of both

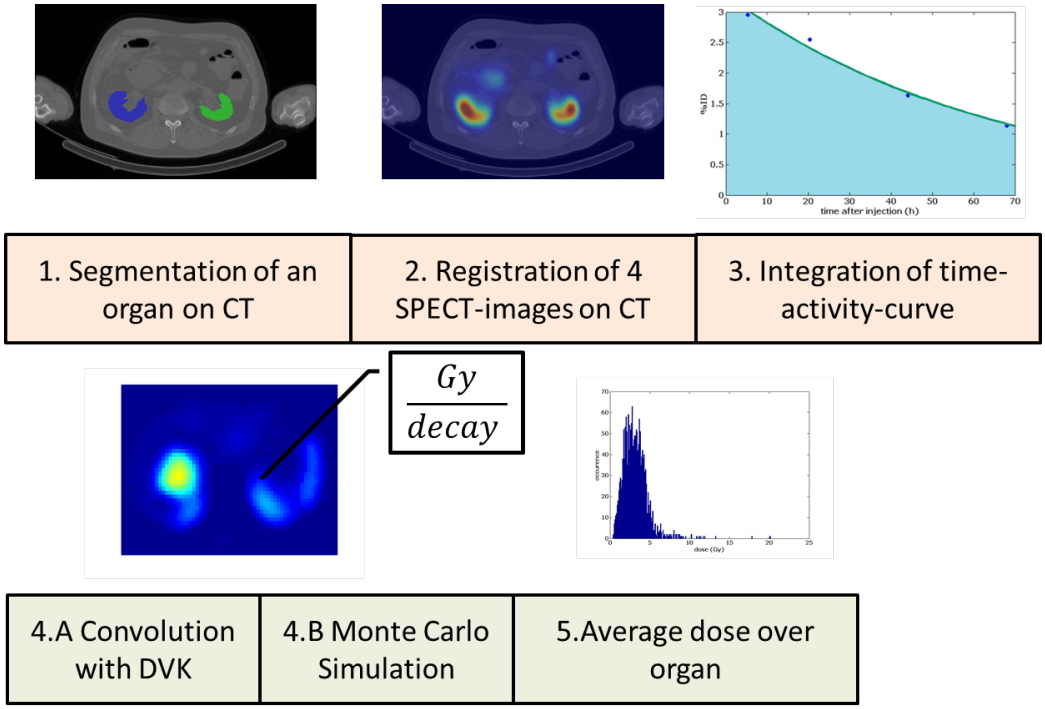


Figure 1.2: Steps of a voxel-wise dosimetry chain

- density pattern (*Computed Tomograpy* (CT)) and
- accumulated activity distribution (SPECT),

available in a voxel geometry, form essential resources for any dose estimation. If SPECT images are available at subsequent time points, the time dependence of the activity distribution, the (*Time Activity Curve* (TAC)), can be modeled. In this study, the number of decays, i. e. the integral over the TAC, has been estimated from 4 SPECT images and one CT image.

X-ray CT provides depth information from multiple projections. Image reconstruction from multiple projections assumes that the activity distribution remains stationary throughout the acquisition. This means that the difference in spatial distribution of nuclear disintegrations is, in the acquired projection images, only a function of the projection view. The task for any reconstruction method is then to estimate the individual activity distribution in the patient, possibly at a number of subsequent time points. For many years, *Filtered Back-Projection* (FBP) was the golden standard of reconstruction methods, but today, the clinical use of iterative reconstruction methods is established. This family of reconstruction methods all include a computer model of the imaging system. The principal steps in an iterative reconstruction method are the following:

- From a first estimate of the internal source distribution, the computer model calculates a projection image.
- This image is then compared with a measured projection to determine where in the modeled image differences occur.
- Most reconstructions in nuclear medicine use the *Expectation Maximization* (EM) algorithm [34], [101], [79].

Iterations go on until the ratio between the modeled and the measured data converges to one. The initial image is multiplied by a tomographic error image. The latter is obtained from back-projecting the ratios. This is done for all projection angles. Note that the noise in these images shows a Poisson distribution. A version of the EM algorithm is the ordered-subsets EM algorithm [63].

1.2.1 Time-Activity-Curve (TAC)

The amount and rates of radiopharmaceutical uptake and excretion is governed by passive as well as active physiological mechanisms and varies amongst individuals. These parameters directly influence absorbed dose calculations and need to be determined for each patient. The method used to determine TACs for different organs and tissues is

- to perform imaging at several time points after radiopharmaceutical administration and
- to determine the activity concentration in different tissues at every time point.

The absorbed dose is associated with the area under this TAC. The activity $A(t)$ of a radionuclide is given as the number of nuclear disintegrations $N(t)$ per unit time and decreases exponentially in time. Thus a sum of exponential curves with different time constants can then be used to fit patient data [148] according to

$$A(t) = \sum_{S \in \mathcal{S}} A(\mathbf{r}_S, 0) \exp(-\lambda_{phys} \cdot t)$$

where $A(t) = dN(t)/dt$ denotes the total activity, $A(\mathbf{r}_S, t) = dN(\mathbf{r}_S, t)/dt$ the nuclide-specific activity at voxel location $\mathbf{r}_S$ in the *Volume Of Interest* (VOI) $\mathcal{S}$ and where the rate constant $\lambda_{phys} = 1/(\tau_{phys})$ describes the kinetics of removal of activity components $A(\mathbf{r}_S, t)$ with a nuclide-specific lifetime τ_{phys}. The latter is related to the half-time of the nuclide via $t_{1/2} = \ln(2)\tau$. In addition to this

physical lifetime, also a biological lifetime is of relevance for removing activity components. If the biological removal also follows an exponential law, the rate constants can be summed up to an effective rate constant

$$\lambda_{eff} = \lambda_{phy} + \lambda_{biol} \tag{1.1}$$

which results in an effective lifetime

$$\tau_{eff} = \frac{\tau_{biol} \cdot \tau_{phys}}{\tau_{biol} + \tau_{phys}} \tag{1.2}$$

This finally results in a *Time Integrated Activity* (TIA) given by

$$\begin{aligned} \tilde{A}(\mathbf{r}_S) \equiv N(\mathbf{r}_S) &= \int_0^\infty A(t)dt = \sum_i \int_0^\infty A(\mathbf{r}_S, 0) e^{-\lambda_{eff}(\mathbf{r}_S)\cdot t} dt \\ &= \sum_S \frac{A(\mathbf{r}_S, 0)}{\lambda_{eff}(\mathbf{r}_S)} = \sum_S \cdot A(\mathbf{r}_S, 0) \cdot \tau_{eff}(\mathbf{r}_S) \end{aligned} \tag{1.3}$$

where the sum extends over all voxels of the entire VOI. Note that for the effective lifetime we have

$$\tau_{eff} \leq \begin{cases} \tau_{phys} & \text{if} \quad \tau_{phys} \leq \tau_{biol} \\ \tau_{biol} & \text{if} \quad \tau_{biol} < \tau_{phys} \end{cases} \tag{1.4}$$

Note also that solving the equation for an effective lifetime, the units for the biological and physical lifetimes must be the same. If a voxel-wise dosimetry is performed, the integration of the time activity curve results in a voxel-wise map of the number of decays per voxel. For the MIRD method, the number of decays, which take place in the kidney, can be calculated by integration.

1.2.2 Energy dose rate and energy dose

The rate $dD_E(\mathbf{r}_T, t)/dt$ at which, at time t after administration, radiation energy dose is delivered within a patient to target tissue at location $\mathbf{r}_T$ from a radioactive material distributed *uniformly* within source tissue located at $\mathbf{r}_S$ is given as:

$$\frac{dD_E(\mathbf{r}_T, t)}{dt} := \dot{D}_E(\mathbf{r}_T, t) = A(\mathbf{r}_S, t) K(\mathbf{r}_T \leftarrow \mathbf{r}_S, t) \left[\frac{J}{kg \cdot s} = \frac{Gy}{s}\right] \tag{1.5}$$

where $A(\mathbf{r}_S, t)$ is the time-dependent activity of a radionuclide in source tissue $\mathbf{r}_S$, i.e. $A(\mathbf{r}_S, t)$ counts the number of nuclear disintegrations happening at time t in source tissue located at $\mathbf{r}_S$. Further, $K(\mathbf{r}_T \leftarrow \mathbf{r}_S, t) = \dot{D}_E(\mathbf{r}_T, t)/A(\mathbf{r}_S, t)$ denotes

the radionuclide-specific energy dose kernel representing, at time t after administration, the energy dose absorbed in target tissue $\mathbf{r}_T$ per unit activity present in source tissue $\mathbf{r}_S$. The *energy dose rate*, i. e. $d^2E(\mathbf{r}_T,t)/dm\ dt = dD_E/dt$, thus denotes how much energy is deposited in a unit mass in unit time. In practice, more handy units are $\frac{\mu Gy}{hr}$ or $\frac{mGy}{yr}$.

Absorbed energy dose $D_E(\mathbf{r}_T)$, i. e. the deposited radiation energy per unit mass over a timespan T_D, is the fundamental quantity for coupling to the radiobiological effect. It is defined as the radiation energy $dE(\mathbf{r}_T)$ absorbed in a target tissue contained in a volume $dV_T = d^3r_T$ with mass density ρ_m yielding a target mass $dm = \rho_m d^3 r_T$, i. e. we have

$$\begin{aligned} D_E(\mathbf{r}_T) &:= \frac{dE(\mathbf{r}_T)}{dm} \\ &= \frac{dE(\mathbf{r}_T)}{\rho_m dV_T} = \frac{\epsilon(\mathbf{r}_T)}{\rho_m} \\ &= \int_0^{T_D} \frac{dD_E(\mathbf{r}_T,t)}{dt} dt \\ &= \sum_{r_S} \int_0^{T_D} A(\mathbf{r}_S,t) K(\mathbf{r}_T \leftarrow \mathbf{r}_S,t) dt \\ &\approx \sum_{r_S} \tilde{A}(\mathbf{r}_S) K(\mathbf{r}_T \leftarrow \mathbf{r}_S) \end{aligned} \tag{1.6}$$

measured in $\left[\frac{J}{kg} \equiv Gy\right]$. The time-dependent activity in the source tissue $A(\mathbf{r}_S,t)$ is obtained

- by numeric solution of a set of first-order coupled differential equations defined by compartment models for all organs and suborgan tissues of interest.
- directly via quantitative imaging, including planar imaging, SPECT, and *Positron-Emission Tomography* (PET), or by tissue sampling (e.g., biopsy, blood, or urine collection).

In most instances, the time dependence of the dose kernel $K(\mathbf{r}_T \leftarrow \mathbf{r}_S,t)$ may be neglected, as when the source and target masses remain constant over the period of irradiation, justifying the use of the approximation given with the last equation. The integral

$$\tilde{A}(\mathbf{r}_S) = \int_{t_1}^{t_2} A(\mathbf{r}_s,t) dt \tag{1.7}$$

is usually called *time-integrated* or *cumulated* activity (TIA) and describes the total number of decays occurring in the source region during the time interval T_D.

In nuclear medicine, due to tissue heterogeneities, *mean absorbed energy dose* $\langle D_E(\mathbf{r}_T)\rangle$ is quoted, which is given by

$$\langle D_E(\mathbf{r}_T)\rangle = \frac{1}{M_T}\int_{V_T} D_E(\mathbf{r})\rho_m(\mathbf{r})d^3r \tag{1.8}$$

where M_T is the total mass of tissue inside a target volume V_T. The latter might represent a tissue voxel or a whole organ for which the absorbed dose is determined.

In order to account for the biological effect of the radiation, equivalent energy dose $D_{eq} = w \cdot D_E\ [Sv]$ has to be considered, where Sv denotes *Sievert* and w represents a tissue-related weight factor (often called relative biological efficiency). The weights are $w = 1$ for electrons and photons, but $w = 20$ for α - particles and $w = 2 - 20$ for neutrons depending on their kinetic energy.

If the *fluency* Ψ of a particle beam is known, the energy dose can be computed as

$$\begin{aligned} D_E &= N\frac{dE}{dm} = N\frac{dE}{\rho_m dV} = \frac{1}{\rho_m}\frac{N}{\sigma}\frac{dE}{dx} \\ D_E \cdot \rho_m &= \Psi \cdot LET \end{aligned}$$

where N is the number of particles, σ the beam cross section, $\Psi = N/\sigma$ is the fluency, ρ_m the mass density and dE/dx the energy loss per unit length, also called stopping power or linear energy transfer.

Energy dose should not be confused with radiation dose D_Q, which denotes the amount of charge $dQ\ [C]$ deposited in a unit mass $dm\ [kg]$. Accordingly we have

$$D_Q = \frac{dQ}{dm}\ \left[\frac{C}{kg}\right]$$

1.2.3 Radiation energy transport and dose kernels

In Radiotherapy, any dosimetry of internally distributed radionuclides relies on the description of the radiation energy transport from a source to a target region [160], i. e.

$$\langle D_E(\mathbf{r}_T)\rangle = \sum_{r_S} \tilde{A}(\mathbf{r}_S) \cdot K(\mathbf{r}_T \leftarrow \mathbf{r}_S)$$

Here the dose kernel is given by

$$K(\mathbf{r}_T \leftarrow \mathbf{r}_S) = \frac{c}{M_T} \sum_i n_i E_i \Phi\left(\mathbf{r}_T \leftarrow \mathbf{r}_S, E_i\right)$$

where $c = 1\,[(Gy{\cdot}kg)/(MBq{\cdot}s{\cdot}MeV)]$ if $D_E\,[Gy], \tilde{A}\,[MBq{\cdot}s], E_i\,[MeV], m\,[kg]$, and $c = 2.13$ if $D_E\ [rad], A\ [\mu Ci], m\ [g], E_i\ [MeV]$ [94]. The dose kernel $K(\mathbf{r}_T \leftarrow \mathbf{r}_S)$, if summed up over all fractions of absorbed energy $\Phi(\mathbf{r}_T \leftarrow \mathbf{r}_S, E_i)$ within a whole organ, is commonly called S-value. It is measured in $[Gy/(Bq{\cdot}s])$. Here the absorbed fraction $\Phi\left(\mathbf{r}_T \leftarrow \mathbf{r}_S, E_i\right)$ of incident radiation energy represents the ratio of the energy E absorbed in target region $\mathbf{r}_T$ to the energy E_0 emitted in the source region $\mathbf{r}_S$

$$\Phi(\mathbf{r}_T \leftarrow \mathbf{r}_S) = \frac{E(\mathbf{r}_T)}{E_0(\mathbf{r}_S)}$$

Furthermore, specific absorbed fraction Φ_{spec} of energy is the ratio of absorbed fraction by the mass of the target tissue m_T

$$\Phi_{spec}(\mathbf{r}_T \leftarrow \mathbf{r}_S) = \frac{\Phi(\mathbf{r}_T \leftarrow \mathbf{r}_S)}{m_T}$$

Thus the S-value represents the *mean absorbed dose* $\langle D_E(\mathbf{r}_T)\rangle$ to the target organ centered at location $\mathbf{r}_T$ per *unit of accumulated activity* $\tilde{A}(\mathbf{r}_S)$ in the source region $\mathbf{r}_S$, that is to say, it describes the fraction of the kinetic energy $E_{kin}\ [J]$ released by each particle type, i, with n_i emitted particles per radioactive decay, which will be absorbed in the target volume with total mass $M_T(t)\ [kg]$. Hereby $\Delta_i = n_i E_i$, where E_i represents the energy emitted for radiation type i with probability n_i. In principle, the equation is valid for different kinds of volumes, from organs down to cells, but for smaller volumes, the stochastic nature of radiation interaction renders the mean value less representative of the actual energy deposition.

1.2.4 MIRD-method

The method defines a protocol for an estimation of the absorbed energy dose to organs of interest using anatomic phantoms representative of any patient or, at least, an average patient. This method computes the *mean organ dose* due to

- electron and α-particle emissions with their accompanying photon emissions within the source organ, and
- the energy dose deposited by photons emitted within surrounding organs and tissues.

This technique represents a macroscopic approach to dose estimation and is currently implemented in the *Organ Level Internal Dose Assessment with Exponential Modeling* (OLINDA) software [149]. The method is easy to use, but only an *average absorbed dose* $\langle D_E \rangle$ can be estimated. Neither heterogeneities in either tissue composition or radioactivity distribution nor the anatomic geometry of the source and target tissues within the body of the patient are considered.

Dosimetric parameters relevant to the MIRD protocol

In any real internal dose problem, there will be more than one organ which concentrates the activity, and many targets for which the absorbed dose is required. In this case, the MIRD equation needs to be solved for each source region $\mathbf{r}_S$ and target region $\mathbf{r}_T$ as follows [150].:

$$\begin{aligned} D_E(\mathbf{r}_T) &= \sum_S \tilde{A}(\mathbf{r}_S) S(\mathbf{r}_T \leftarrow \mathbf{r}_S) \\ &= A_0 \sum_S \tau_S S(\mathbf{r}_T \leftarrow \mathbf{r}_S) \\ &= \frac{c}{M_T} \sum_S w_S \tilde{A}(\mathbf{r}_S) \sum_i n_i E_i \Phi_i(\mathbf{r}_T \leftarrow \mathbf{r}_S) \end{aligned}$$

where a source *residence time* τ_S is defined as

$$\tau_S = \frac{\tilde{A}(\mathbf{r}_S)}{A_0}$$

and where A_0 is the *administered activity*. Thus more than one organ can be taken into consideration using compartmental modeling [107]. Goodness of fit can be assessed through an Akaike information criterion (AIC) [81] or an F-test [89]. Attempts have also been made to compile patient data in order to use the group behavior as a prior when applying Bayesian techniques.

These generic expressions comply well with the most recent MIRD expressions discussed above. The OLINDA software uses a somewhat simpler expression given by

$$\begin{aligned} D &= N \cdot DF \\ DF &= \frac{k}{m} \sum_i n_i E_i \Phi_i w_{r_i} \end{aligned}$$

where DF is conceptually similar to the S-value defined in the MIRD system. The number $N \equiv \tilde{A}(\mathbf{r}_S)$ of disintegrations is the integral of a time-activity curve

for a source region. A number of anthropomorphic phantoms exist that can be used to produce dose estimates for standardized individuals.

1.2.5 Voxelwise dosimetry

Voxelwise dosimetry represents a mesoscopic approach to dose estimation. It means employing voxelized geometries in representing organs as tagged voxels of differing activity levels. MC radiation transport simulations then assess dose at the voxel level. Geometries have to be deduced from phantoms or from a regional CT image of the patient [144], [171]. The MC approach, furthermore, can yield a Dose Voxel Histogram for radiation sources both inside and outside the target organ.

The average energy dose $\langle D_E \rangle$ absorbed at any target organ can be obtained by tracking energy deposition events within the target organ by direct MC transport simulations. The computations consider two contributions:

- The average absorbed dose to a voxel within the organ of interest due to α- and β-particle as well as γ-photon emissions within that organ
- The energy dose to a voxel in the organ of interest due to γ-photon emissions in the surrounding organs or tissues.

The technique can handle tissue heterogeneities (bone, soft tissue, air, lung, and such) as well as patient-specific anatomic geometry of all source and target tissues as well as the non-uniform source distributions [61], [60], [35], [123]. Because of statistical sampling limitations, the energy dose is usually averaged across a group of voxels or all voxels of a given target organ. However, the approach is time-consuming and computationally demanding. For anatomic regions characterized by a uniform tissue density, the S-value technique represents an excellent option, with rapid computation and still-good dosimetric accuracy [36]. In particular, the S-value concept is possibly the most applied, easy to implement, not requiring volume and complying with the familiar MIRD formalism [84]. But the S-value method represents a macroscopic approach lacking resolution at the voxel level. With it, dose values can only be calculated for whole organs.

If one is to boil down the resolution to the size of single voxels, currently two different methods exist for performing internal dosimetry at the voxel level [157], [152]:

- the above mentioned full Monte Carlo simulation (MC) or
- a convolution with a dose-voxel-kernel (DVK).

When the amount and distribution of radioactivity $A(\mathbf{r}_T, t)$, represented in the SPECT- or PET-image is known, the resulting energy dose distribution $D_E(\mathbf{r}_T, t)$ can be determined by transforming the activity distribution from the image to absorbed radiation energy using an appropriate dose kernel $K(\mathbf{r}_T \leftarrow \mathbf{r}_S)$. The source and target regions centered at $\mathbf{r}_S$ and $\mathbf{r}_T$, respectively, are those defined within the anatomic model and may represent voxels from SPECT or PET images. Hence, $\mathbf{r}_S$ denotes the location of the source voxel and $\mathbf{r}_T$ the corresponding location of the target voxel. The kernel thus contains information on how the radiation energy resulting from one nuclear disintegration is spatially deposited. The kernel transfers an activity distribution to an absorbed energy distribution by weighting the energy deposition of the particles over its range. An energy deposition image is obtained by applying the kernel to the original activity distribution image through a convolutive filtering process. Each voxel assigns energy, measured in either $[J]$ or $[MeV]$, and by dividing it with the mass of the voxel, a specific energy, or absorbed dose, measured in $[Gy]$ can be generated in each voxel of the image. The values in this specific energy (or dose) image can then be used to generate a Dose-Volume Histogram [40]. The latter presents the volume (or a percentage of the volume) that receives some dose in a given range of dose levels. If the respective volume refers to one voxel, this is referred to as DVK dosimetry. The DVK method is widely used today, particularly for the more commonly used radionuclides which emit electrons, including Auger electrons, and photons.

In summary, at the voxel level, the absorbed dose to a target voxel located at $\mathbf{r}_T$ due to the cumulated activities in source voxels located at $\mathbf{r}_S$ can be expressed as:

$$D_E(\mathbf{r}_T) = \sum_{S \in \mathcal{S}} \tilde{A}(\mathbf{r}_S) K_{vox}(\mathbf{r}_T \leftarrow \mathbf{r}_S)$$

where the sum is extended to all source voxels. In voxel-wise dosimetry, the DVK is defined as the absorbed dose to the target voxel per unit decay in the source voxel, when both voxels are contained in an infinite homogeneous medium of mass density ρ_m. According to this definition, DVKs can be obtained by MC simulations of voxel geometries representing an infinite medium of uniform composition. The radionuclide is placed at the center of this geometry, and the absorbed dose is scored in surrounding voxels. The mathematical convolution of pre-calculated DVKs, i. e. energy dose per unit decay as a function of distance, with a cumulated activity map provides the absorbed dose distribution for the biological system considered [98].

1.3 Uncertainties in dose estimation

Any error analysis has to start with the data collection systems. Data uncertainty my result from limitations on energy resolution, low spatial resolution due to collimator septal penetration by high energy photons, data loss due to scatter and attenuation as well as inherent statistical variations of measurements on radionuclides. Reports on total error quantifications range from a few percent for large organs to some $20\ \%$ for small or low *Signal to Noise Ratio* (SNR) objects. A summary of the terms contributing to uncertainties in calculated dose estimates in nuclear medicine was given by [150] in terms of the variability of the nuclear medicine population. The conclusion was that the two variables with highest impact onto error estimates were the cumulated activity $\tilde{A}$ and organ masses m. Further issues to be considered are the following:

- The biokinetic parameters, fractional uptake and half-life, vary substantially (factor > 2) across individuals. So, just applying a dose coefficient, measured in $[mSv/MBq]$, from a standard model to any given patient leads to a relatively large uncertainty in dose estimation [106]. It is unfortunate that patient-individualized dose calculations are not performed routinely for therapeutic administrations of radiopharmaceuticals, although this practice is increasing in Europe [44], [11], [19], [117].

- Simply administering a standard activity ($\tilde{A}\ [MBq]$) or specific activity ($\tilde{A}/M\ [MBq/kg]$, $\tilde{A}/O\ [MBq/m^2]$) per unit body mass M or surface area O to all patients cannot permit the delivery of an adequate therapy to all patients. In general, several nuclear medicine images are needed, yielding $A(\mathbf{r}_S, t)$ to establish the uptake and clearance patterns in normal tissues and tumors [147].

Chapter 2

Monte Carlo Simulation (MC)

The following short summary lends credit from [75].

Photons interact with surrounding matter via four basic processes:

- *Rayleigh scattering*: A coherent scattering with the molecules (or atoms) of the medium
- *Photoeffect*: A photo-electric absorption of a photon and transfer of its energy to an electron
- *Compton scattering*: An incoherent scattering with atomic electrons
- *Particle production*: A materialization into an electron/positron pair in the electromagnetic field of the nuclei and surrounding atomic electrons

The last three collision types transfer energy from the photon beam to electrons. Depending on energy and the medium, in which the transport takes place, one of them dominates the total absorption cross-section:

- At low energies, $E < 100\ keV$, the *Photoeffect* dominates .
- At intermediate energies, $100\ keV \leq E \leq 1\ MeV$, the *Compton effect* is the most important process .
- The *pair production* process dominates at energies $E \geq 1.02\ MeV$ larger than the mass - equivalent energy of the resting particle pair.

Electrons, as they traverse matter, loose energy by three basic processes:

- Inelastic collisions with atomic electrons
- Bremsstrahlung and positron annihilation

Radiative energy loss transfers energy back to photons and leads to the coupling of the electron and photon radiation fields. The bremsstrahlung process is the dominant mechanism of electron energy loss at high energies, inelastic collisions are more important at low energies. In addition, electrons participate in elastic collisions with atomic nuclei which occur at a high rate and lead to frequent changes in the electron direction. Inelastic electron collisions and photon interactions with atomic electrons lead to excitations and ionizations of the atoms along the paths of the particles. Highly excited atoms, with vacancies in inner shells, relax via the emission of photons and electrons with characteristic energies.

The coupled integro-differential equations that describe the electromagnetic shower development are prohibitively complicated to allow for an analytical treatment except under severe approximations. The MC technique is the only known solution method that can be applied for any energy range of interest. MC of particle transport processes are a faithful simulation of physical reality:

- Particles are drawn from a source distribution deduced from activity distributions known across a target VOI
- These particles travel certain distances, determined by a probability distribution depending on the total interaction cross section Ω
- At the site of a collision, they scatter into another energy and/or direction according to the corresponding differential cross section σ
- During such collisions, possibly new particles are produced that have to be transported as well.

This procedure is continued until all particles are absorbed or leave the VOI under consideration.

Quantities of interest can be calculated by averaging over a given set of *MC particle histories* (also refereed to as showers or cases). From mathematical points of view each particle history is one point in a d-dimensional space (the dimensionality depends on the number of interactions) and the averaging procedure corresponds to a d-dimensional MC integration. As such, the MC estimate of quantities of interest is subject to a *statistical uncertainty* which depends on N, the number of particle histories simulated, and usually decreases as $N^{-1/2}$. Depending on the problem under investigation and the desired statistical accuracy, very long calculation times may be necessary.

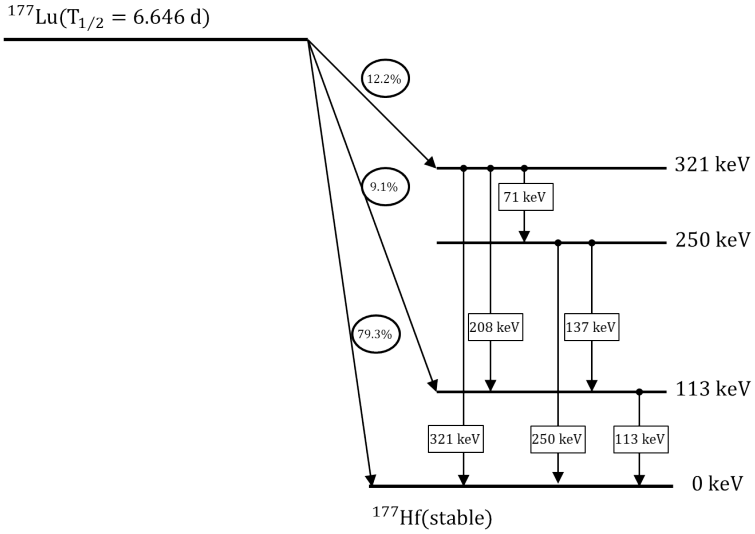
An additional difficulty occurs in the case of the MC of electron transport. In the process of slowing down, a typical fast electron and the secondary particles it creates undergo hundreds of thousands of interactions with surrounding matter.

Because of this large number of collisions, an event-by-event simulation of electron transport is often not possible due to limitations in computing power. To circumvent this difficulty, Berger [14] developed the "condensed history" technique for the simulation of charged particle transport. In this method, large numbers of subsequent transport and collision processes are "condensed" to a single "step". The cumulative effect of the individual interactions is taken into account by sampling the change of the particle's energy, direction of motion, and position, at the end of the step from appropriate multiple scattering distributions. The condensed history technique was motivated by the fact that single collisions with the atoms cause in most cases only minor changes in the particle's energy and direction of flight. This technique made MC of a charged particle transport possible but introduced an artificial parameter, the step-length. The dependence of the calculated result on the step-length has become known as a *step-size artifact* [17].

2.1 Radionuclide

For estimating the absorbed radiation dose elicited from a ^{177}Lu radiation source into the surrounding tissue (air, lung, bone), N MC have been performed, each encompassing M nuclear disintegrations, shortly called radioactive decays.

During one such decay, ^{177}Lu transforms into ^{177}Hf according to the following state transitions.



The related γ - radiation shows several spectral lines according to the scheme seen above and the lines exhibit very small line widths. Hence, one could model the spectrum as a sum of discrete Lorentzians.

Here a Lorentzian line is given in the time domain and its conjugated Fourier domain, respectively by

$$\begin{aligned} L(t) &= H(t)L_0 \exp(-t/\tau) \\ L(\omega) &= \frac{L_0}{1/\tau + i\omega} = \frac{L_0}{\sqrt{(1/\tau)^2 + \omega^2}} \exp(-i\phi) \end{aligned}$$

where $H(t)$ denotes the Heaviside step function, $\omega = 2\pi\nu$ with $\nu\ [Hz]$ the frequency, τ the average life-time, $\phi = \arctan(\omega\tau)$ and $i = \sqrt{-1}$ denotes Euler's number. The Lorentzian line in the frequency domain consist of a real and an imaginary part according to

$$\begin{aligned} L(\omega) &= \frac{(1/\tau)L_0}{(1/\tau)^2 + \omega^2} - i\frac{\omega L_0}{(1/\tau)^2 + \omega^2} \\ &= \tau L_0 \left(\frac{1}{1 + (\omega\tau)^2} - i\frac{\omega\tau}{1 + (\omega\tau)^2} \right) \end{aligned}$$

The real part of this complex Lorentzian provides the conventionally observed absorption mode, often called a Lorentz line.
Hence, using ^{177}Lu the main dates are:

Table 2.1: Relevant characteristic quantities of a ^{177}Lu radiation

Nuclide	$\beta(MeV)$	Range (mm)	$T_{1/2}\ (d)$	$\gamma(keV)$
^{177}Lu	$0,50$	2	$6,7$	$113(6\%)$

2.2 Standard Error for Dose Estimation

In the following, the dose will be denoted by X and is considered a random variable resulting from independent nuclear disintegrations, simply called decays. Hence, x_{mn} denotes the decay event m happening in simulation run n inside a voxel of volume $V_{vox}\ [mm^3] = 4.7\ [mm] \cdot 4.7\ [mm] \cdot 4.7\ [mm]$ and tissue density $\rho_m\ [mg \cdot mm^{-3}]$ where m denotes the average mass of either air, lung tissue or bone.

Following, we will consider two strategies:

- Consider M simulated decays as one sample of data and simulate N such samples in total

- Consider NM simulated decays as one grand sample of data and simulate L such grand samples in total

Typically we talk about $N \approx 10^2$ and $M \approx 10^4$.

Relation between standard error (SE) and standard deviation (SD) The *Standard Error* (SE) denotes the *Standard Deviation* (SD) of a sampling distribution of a statistic. The latter can be, for example, the average energy dose $\langle D_E(\mathbf{r}_T)\rangle_{V_{vox}}$ in a voxel of tissue. Hence, the SE depicts the dispersion of sample means around the population mean, while the SD denotes the dispersion of individual measurements around the population mean.

Now for a given sample size, the SE equals the SD divided by the square root of the *Sample Size* (SS), i. e.

$$SE = \frac{SD}{\sqrt{SS}}$$

Following we denote the sample SE by se_n and the grand SE by SE, the sample SD by sd_n and the grand SD by SD.

Sample SE First we consider each ^{177}Lu decay as an independent event and we simulate M such decays. We then consider these M simulated decays as one sample of data. Each such decay deposits an amount x_{mn} $[Gy]$ of energy dose in the surrounding tissue. Averaging over M independent decay events yields an average energy dose

$$\langle x_n \rangle = \frac{1}{M} \sum_{m=1}^{M} x_{mn}$$

which we call *sample mean* and which is deposited in one voxel during one single simulation run n. The size of the sample is now $SS = M$. Each single dose x_{nm}, deposited in one voxel during one decay event, is scattered around its sample mean $\langle x_n \rangle$, and this scatter is characterized by the corrected *sample standard deviation*

$$sd_n = \sqrt{\frac{1}{M-1} \sum_{m=1}^{M} \left(x_{mn} - \langle x_n \rangle\right)^2}$$

where the factor in front of the sum term represents the Bessel correction ($M-1$ instead of M), which assures an unbiased variance (not standard deviation! The latter is still biased due to the non-linear sqrt operation). The sample standard deviation then yields a sample standard error according to

$$
\begin{aligned}
se_n &= \frac{sd_n}{\sqrt{M}} \\
&= \frac{1}{\sqrt{M}} \sqrt{\frac{1}{M-1} \sum_{m=1}^{M} \left(x_{mn} - \langle x_n \rangle\right)^2} \\
&\approx \frac{1}{M} \sqrt{\sum_{m=1}^{M} \left(x_{mn} - \langle x_n \rangle\right)^2} \\
&\approx \frac{1}{M} \sqrt{\sum_{m=1}^{M} x_{mn}^2 - M \langle x_n \rangle^2}
\end{aligned} \tag{2.1}
$$

Repeating such simulations N times, yields N samples of mean energy doses deposited in a voxel. These N samples provide an estimate of the underlying unknown *sampling distribution* of mean energy doses $\langle x_n \rangle$ deposited in each voxel.

The underlying sampling distribution can be characterized by

- a *grand average* deposited energy dose which, in abuse of notation, could be called a population mean

$$\langle\langle x \rangle\rangle = \frac{1}{N} \sum_{n=1}^{N} \langle x_n \rangle = \frac{1}{MN} \sum_{n=1}^{N} \sum_{m=1}^{M} x_{mn}$$

  Note that if $N \to \infty$, then the grand average tends towards the population mean energy dose.

- the *grand mean standard deviation* of the fluctuating sample means around the grand average, i. e. the width of the sampling distribution, which we could accordingly call the standard deviation of the mean

$$
\begin{aligned}
SD &= \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} \left(\langle x_n \rangle_M - \langle\langle x \rangle\rangle_{MN}\right)^2} \\
&= \sqrt{\frac{1}{N-1} \left[\sum_{n=1}^{N} \langle x_n \rangle_M^2 - N \langle\langle x \rangle\rangle_{MN}^2\right]}
\end{aligned}
$$

As we have estimates of all quantities available entering the last line, we can easily estimate the SD of the sampling distribution.

The related standard error SE is then obtained via

$$\begin{aligned} SE &= \frac{SD}{\sqrt{N}} \\ &= \frac{1}{\sqrt{N}}\sqrt{\frac{1}{N-1}\left[\sum_{n=1}^{N}\langle x_n\rangle_M^2 - N\langle\langle x\rangle\rangle_{MN}^2\right]} \\ &\approx \frac{1}{N}\sqrt{\sum_{n=1}^{N}\langle x_n\rangle_M^2 - N\langle\langle x\rangle\rangle_{MN}^2} \\ &\approx \frac{1}{\sqrt{N}}\sqrt{\langle\langle x_n\rangle_M^2\rangle_N - \langle\langle x\rangle\rangle_{MN}^2} \end{aligned} \tag{2.2}$$

Substituting from above yields

$$\begin{aligned} SE &= \frac{SD}{\sqrt{N}} \\ &= \frac{1}{\sqrt{N}}\sqrt{\frac{1}{N-1}\left[\sum_{n=1}^{N}\left(\frac{1}{M}\sum_{m=1}^{M}x_{mn}\right)^2 - N\left(\frac{1}{MN}\sum_{n=1}^{N}\sum_{m=1}^{M}x_{mn}\right)^2\right]} \\ &\approx \frac{1}{MN}\sqrt{\sum_{n=1}^{N}\left(\sum_{m=1}^{M}x_{mn}\right)^2 - N\left(\sum_{n=1}^{N}\sum_{m=1}^{M}x_{mn}\right)^2} \end{aligned} \tag{2.3}$$

Relative SE A relative SE can be obtained by simply dividing SE by the mean according to

$$\begin{aligned} rSE &= \frac{SE}{\langle\langle x\rangle\rangle}\cdot 100 \\ &\approx \frac{100}{\sqrt{N}}\sqrt{\frac{\langle\langle x_n\rangle_M^2\rangle_N - \langle\langle x\rangle\rangle_{MN}^2}{\langle\langle x\rangle\rangle_{MN}^2}} \end{aligned} \tag{2.4}$$

$$\approx 100\cdot\sqrt{\frac{\sum_{n=1}^{N}\left(\sum_{m=1}^{M}x_{mn}\right)^2}{\left(\sum_{n=1}^{N}\sum_{m=1}^{M}x_{mn}\right)^2} - N} \tag{2.5}$$

Total SE Alternatively, this collection of N simulations, each consisting of M decays, can be considered one *sample of data* of size MN. Sample mean energy dose and related sample standard deviation and sample standard error can then be computed as follows:

$$\begin{aligned} tSE &= \frac{1}{\sqrt{NM}}\sqrt{Var(x_{nm})} = \\ &= \frac{1}{\sqrt{NM}}\sqrt{\frac{1}{NM-1}\left[\sum_n^N\sum_m^M(x_{nm}-\langle\langle x\rangle\rangle)^2\right]} \\ &= \frac{1}{\sqrt{NM}}\sqrt{\frac{1}{NM-1}\left[(\sum_n^N\sum_m^M x_{nm}^2)-NM\cdot\langle\langle x\rangle\rangle^2\right]} \\ &\approx \frac{1}{\sqrt{NM}}\sqrt{[\langle\langle x^2\rangle\rangle_{MN}-\langle\langle x\rangle\rangle_{MN}^2]} \end{aligned} \tag{2.6}$$

$\langle\langle x\rangle\rangle$ is the mean over all mean values from each run.

The values for x_{nm} are unknown, thus the term $\sum_m^M x_{nm}^2$ has to be replaced by an expression of known quantities. To achieve this goal, consider the standard error for a single run:

$$\begin{aligned} se_n &= \frac{1}{M}\sqrt{\frac{1}{M-1}[(\sum_m^M x_{nm}^2)M - M^2\cdot\langle x\rangle_n^2]} \\ \sum_m^M x_{nm}^2 &= se_n^2M(M-1)+M\cdot\langle x\rangle_n^2 \end{aligned}$$

Because $M \gg 1$ this can be easily approximated by the following expression

$$\begin{aligned} \frac{1}{M}\sum_m^M x_{nm}^2 &= se_n^2(M-1)+\langle x\rangle_n^2 \\ \langle x_n^2\rangle &= se_n^2(M-1)+\langle x\rangle_n^2 \\ \langle x_n^2\rangle &\approx M\cdot se_n^2+\langle x\rangle_n^2 \end{aligned}$$

The term $\sum_m^M x_{nm}^2$ can be replaced now by the exact expression, and the *total standard error* (tSE) for the total number of decay events across all sessions, i. e. for all samples, can be calculated as:

$$tSE = \frac{1}{NM}\sqrt{\frac{NM}{NM-1}\left(\sum_n^N((M-1)\cdot se_n^2 + (\langle x\rangle_n)^2) - (NM)\cdot\langle\langle x\rangle\rangle^2\right)}$$

If, instead, the approximate expression is plugged in, one obtains

$$\begin{aligned} tSE &= \frac{1}{NM}\sqrt{\frac{1}{NM-1}\left(\sum_n^N(M\cdot se_n^2 + \langle x\rangle_n^2)N - (NM)^2\cdot\langle\langle x\rangle\rangle^2\right)} \\ &\approx \frac{1}{NM}\sqrt{\sum_n^N(M\cdot se_n^2 + \langle x\rangle_n^2) - (NM)\cdot\langle\langle x\rangle\rangle^2} \end{aligned}$$

Relative SE: We are interested in the *relative standard error*. Therefore, we have to divide the tSE by the population mean $\langle\langle x\rangle\rangle$ and replace the se_n.

$$\begin{aligned} rtSE = \frac{1}{NM}\frac{1}{\langle\langle x\rangle\rangle}\Bigg[\frac{1}{NM-1}\Bigg(\sum_{n=1}^N\left(rse_n^2\langle x\rangle_n^2 M(M-1) + M\cdot\langle x\rangle_n^2\right)N \\ -\frac{(NM)^2}{1}\cdot\langle\langle x\rangle\rangle^2\Bigg)\Bigg]^{1/2} \end{aligned}$$

However, estimating a grand mean ("population") energy dose and related standard error can be obtained only if a set of samples is collected.

Collecting $L \gg 1$ samples would allow to estimate the underlying sampling distribution, characterized by the grand mean energy dose and its "population" standard deviation. This latter standard deviation divided by the grand mean is called the *total standard error* (total SE tSE), i. e. the SE of the total number of simulations. This could be done in analogy to above using the expressions computed in this section.

Chapter 3

Particle filter (PF)

Non-linear dynamical systems evolve either continuously or via transitions between discrete states. The latter approximation is only reasonable if the mean residence time in any state is longer than the transition time. Particle filters represent discrete state space models which estimate the states $\mathbf{X}(\mathbf{r}, t)$ of a dynamical system from noisy and disturbed observations $\mathbf{Z}(\mathbf{r}, t)$. This is achieved by combining available measured data $\mathbf{Z}(\mathbf{r}, t)$ with prior knowledge about the underlying physical phenomena driving the dyanmics and the measurement process. The underlying physics allow to construct a state evolution model while the known measurement process provides information to construct an observation model. With these ingredients and observations of the system dynamics, estimates of the unobservable dynamic state variables can be produced sequentially.

State evolution models are commonly based on a Markov model, while observation models need to reflect the specifics of the physical measurement process. Both, evolution and observation models are most often described in terms of conditional probability density functions. Hence we have (throughout we use the short hand notation $\mathbf{z}(t_k) \equiv \mathbf{z}_k$, $\{\mathbf{z}(t_0), \ldots, \mathbf{z}(t_k)\} \equiv \mathbf{z}_{0:k}$ etc.)

$$\mathbf{x}_{k+1} \sim p\left(\mathbf{x}_{k+1}|\mathbf{x}_{0:k}\right) \approx p\left(\mathbf{x}_{k+1}|\mathbf{x}_k\right) \tag{3.1}$$

$$\mathbf{z}_k \sim p\left(\mathbf{z}_k|\mathbf{x}_{0:k}\right) \approx p\left(\mathbf{z}_k|\mathbf{x}_k\right) \tag{3.2}$$

where $\sim$ means "represented through", $\mathbf{x}_k$ denotes an unobservable state vector of the dynamical system and $\mathbf{z}_k$ a corresponding observation vector of the system through a measurement at discrete time t_k [96]. Note that the conditional probability densities only depend on the last state of the system. This is in accord with the Markov assumption which expresses common intuition that any prediction of a future state should be based primarily on the recent history of the evolving system.

Now given the task that we want to predict unobservable states of an evolving system from sequential noisy observations, Bayesian inference [111] allows to integrate available information, i. e. observations $\mathbf{z}_{0:k}$ and an associated physical model of the dynamical system, parametrized through dynamic states $\mathbf{x}_{0:k}$. As new information $\mathbf{z}_k$ is available, it can be combined with previous information $p(\mathbf{x}_{0:k-1}, \mathbf{z}_{0:k-1})$ via the Bayes Theorem to deduce the underlying system state $\mathbf{x}_k$. Employing the Markov assumption we have

$$p(\mathbf{x}_k|\mathbf{z}_{0:k}) = p(\mathbf{x}_k|\mathbf{z}_k, \mathbf{z}_{0:k-1}) \approx \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{k-1})}{\sum_k (p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{k-1}))} \tag{3.3}$$

The latter forms the basis for statistical estimates of the posterior density $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ of the unobservable state vector $\mathbf{x}_k$ given the likelihood of the observation $p(\mathbf{z}_k|\mathbf{x}_k)$, some prior knowledge $p(\mathbf{x}_k|\mathbf{z}_{k-1})$ about the observation process and the data evidence $p(\mathbf{z}_k)$.

Note that the term posterior density denotes different densities depending on the application:

- Filtering: $p(\mathbf{x}_k|\mathbf{z}_{0:k})$
- Prediction: $p(\mathbf{x}_{k+\tau}|\mathbf{z}_{0:k})$
- Smoothing: $p(\mathbf{x}_{k-\tau}|\mathbf{z}_{0:k})$

where $\tau > 0$ denotes the prediction or smoothing lag.

The most widely known Bayesian filter method is the Kalman filter [72], which, however, relies on linear models with additive Gaussian noise, i. e. it assumes the distribution $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ to be a multi-variate Gaussian. Extensions of the Kalman filter concern sequential Monte Carlo methods [43] which represent the posterior density in terms of random samples (particles) and associated weights. Such sequential Monte Carlo methods are variously known as *Particle Filter*, bootstrap filter [50], survival of the fittest [73], condensation algorithm [68] and can be applied to non-linear models with non-Gaussian errors [39]. Thus, given any dynamical state representation $f(\mathbf{x}_k)$, one can estimate the related expected value via

$$\begin{aligned} \mathbb{E}(f(\mathbf{x}_k)) &= \int f(\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{0:k})d\mathbf{x}_k \\ &= \frac{\int f(\mathbf{x}_k)p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{k-1})d\mathbf{x}_k}{\int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{k-1})d\mathbf{x}_k} \\ &\approx \sum_{l=1}^{L} w_k^{(l)} f(\mathbf{x}_k^{(l)}) \end{aligned} \tag{3.4}$$

Here $\mathbf{x}_k^{(l)}$ denotes a set of samples, called *particles*, drawn from the posterior distribution $p(\mathbf{x}_k|\mathbf{z}_{k-1})$. The sampling weights are given by

$$w_k^{(l)} = \frac{p\left(\mathbf{z}_k|\mathbf{x}_k^{(l)}\right)}{\sum_{l'=1}^{L} p\left(\mathbf{z}_k|\mathbf{x}_k^{(l')}\right)} \tag{3.5}$$

where the same samples are used in the numerator and the denominator. Hence a particle filter approximates the posterior distribution $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ by a set of particles $\mathbf{x}_k^{(l)}$, sampled from $p(\mathbf{x}_k|\mathbf{z}_{k-1})$, and their weights $\{w_k^{(l)}\}$. The latter are normalized to $0 \leq w_k^{(l)} \leq 1$ and obey the closure relation $\sum_l w_k^{(l)} = 1$.

In practice one is concerned with a sequential sampling scheme, which works as follows:

- Suppose, a set of samples $\{\mathbf{x}_k^{(l)}\}$ and related weights $w_k^{(l)}$ have been obtained at time t_k. These samples and weights represent the posterior distribution $p(\mathbf{x}_k|\mathbf{z}_k)$ at time step t_k.
- Suppose, an observation $\mathbf{z}_{k+1}$ has been made subsequently

How can one estimate corresponding samples $\mathbf{x}_{k+1}^{(l)}$ and weights $w_{k+1}^{(l)}$?

- First, new samples are drawn from the distribution

$$\begin{aligned}
p(\mathbf{x}_{k+1}|\mathbf{z}_{0:k}) &= \int p(\mathbf{x}_{k+1}|\mathbf{x}_k, \mathbf{z}_{0:k}) p(\mathbf{x}_k|\mathbf{z}_{0:k}) d\mathbf{x}_k \\
&= \int p(\mathbf{x}_{k+1}|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_{0:k}) d\mathbf{x}_k \\
&= \int p(\mathbf{x}_{k+1}|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_k, \mathbf{z}_{0:k-1}) d\mathbf{x}_k \\
&= \frac{\int p(\mathbf{x}_{k+1}|\mathbf{x}_k) p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_{k-1}) d\mathbf{x}_k}{\int p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_{k-1}) d\mathbf{x}_k} \\
&= \sum_{l=1}^{L} w_k^{(l)} p\left(\mathbf{x}_{k+1}|\mathbf{x}_k^{(l)}\right)
\end{aligned} \tag{3.6}$$

where the conditional independence property has been used

$$p(\mathbf{x}_{k+1}|\mathbf{x}_k, \mathbf{z}_{0:k}) = p(\mathbf{x}_{k+1}|\mathbf{x}_k) \tag{3.7}$$

$$p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{z}_{0:k-1}) = p(\mathbf{z}_k|\mathbf{x}_k) \tag{3.8}$$

The above distribution is a mixture distribution. Hence, samples $\mathbf{x}_k^{(l)}$ can be drawn by choosing a mixture component l with probability $w_k^{(l)}$ and drawing a sample from the corresponding mixture component of the distribution.

- Second, for each sample the new observation $\mathbf{z}_{k+1}$ is used to estimate the related weights

$$w_k^{(l)} \propto p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}^{(l)}) \tag{3.9}$$

Thus using a probabilistic description and the Markov assumption, one is concerned with the following densities:

- The likelihood of the state $p(\mathbf{x}_k|\mathbf{z}_{k-1})$
- The Markov transition model $p(\mathbf{x}_{k+1}|\mathbf{x}_k) \rightarrow$ state evolution model
- The sensor model $p(\mathbf{z}_k|\mathbf{x}_k) \rightarrow$ observation model

3.1 State Estimation

While theoretical derivations are based on the general Markov model, algorithms are based on evolution and observation models as detailed next.

3.1.1 The model equations

Let the dynamic state of a system be described by a state vector $\mathbf{x}(t) \in \mathcal{R}^N$ whose components represent the dynamical variables of the system. The dynamic evolution of the system, called the state evolution model $\mathbf{f}(...)$, is described by the following equation for the state vector

$$\mathbf{x}(t_k) \equiv \mathbf{x}_k = \mathbf{f}\left(\mathbf{x}_{k-1}, \boldsymbol{\epsilon}^{(x)}\right)$$

where $k = 1, 2, ...$ denotes discrete time instances t_k and $\boldsymbol{\epsilon}^{(x)} \in \mathcal{R}^L$ denotes the state noise.

If measurements are available at t_k, they are collected in the observation vector $\mathbf{z}$, and its relation to the state vector is given by the observation model $\mathbf{h}(...)$

$$\mathbf{z}(t_k) \equiv \mathbf{z}_k = \mathbf{h}\left(\mathbf{x}_k, \boldsymbol{\epsilon}^{(m)}\right)$$

where $\boldsymbol{\epsilon}^{(m)} \in \mathcal{R}^M$ denotes the measurement noise. The various noise distributions $p(\boldsymbol{\epsilon}^{(x)})$ and $p(\boldsymbol{\epsilon}^{(m)})$ are assumed to be known. In some applications, also additional known *control variables* $\mathbf{u}_k$ enter the problem, but are omitted here.

The state estimation problem aims at obtaining information about the state vector $\mathbf{x}_k$ based on the state evolution model $\mathbf{f}(...)$ and based on the measurements $\mathbf{z}_{1:k}$, given by the observation model $\mathbf{h}(...)$.

3.1.2 The simplifying assumptions

The state evolution and observation models rely on the following simplifying assumptions:

- The sequence of state vectors $\mathbf{x}_k$ forms a Markov process [41] and does not depend on the sequence of observations

$$p(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$$

- The observations $\mathbf{z}_k$ form a *Markov process* with respect to the history of the state vector $\mathbf{x}_k$, i. e.

$$p(\mathbf{z}_k|\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}) = p(\mathbf{z}_k|\mathbf{x}_k)$$

- The probability of an observation only depends on the current state

$$p(\mathbf{z}_k|\mathbf{x}_{0:k}, \mathbf{z}_{1:k-1}) = p(\mathbf{z}_k|\mathbf{x}_k)$$

- For $i \neq j$ the noise vectors $\boldsymbol{\epsilon}_i^{(x)}$ and $\boldsymbol{\epsilon}_j^{(x)}$, as well as $\boldsymbol{\epsilon}_i^{(m)}$ and $\boldsymbol{\epsilon}_j^{(m)}$, are mutually independent
- The noise vectors $\boldsymbol{\epsilon}_i^{(x)}$ and $\boldsymbol{\epsilon}_j^{(m)}$ are mutually independent for all $i, j = 1, 2, ...$
- All noise vectors are also mutually independent of the initial state $\mathbf{x}_0$

3.1.3 The problems to be solved

Several problems can be tackled with this approach

- The *prediction problem* determines $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$, i. e. it predicts the state of the system at time t_k based on observations made at times $t_1, \ldots, t_{k-1}$
- The *filtering problem* determines $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, i. e. it filters out from the observations made up to time point t_k the state of the system at time t_k.

- The *fixed lag smoothing problem* determines $p(\mathbf{x}_k|\mathbf{z}_{1:k+\tau})$, i. e. it determines the state of the system at time t_k given observations over a time span $t_1, \ldots, t_{k+\tau}$, where $\tau \geq 1$ denotes the fixed time lag.

- The *whole domain smoothing problem* determines $p(\mathbf{x}_k|\mathbf{z}_{1:K})$, where $\mathbf{z}_{1:K} = \mathbf{z}_i, i = 1, \ldots, K$ denotes the complete observation sequence.

In this study, only the *filtering problem* is of interest here. By assuming that $p(\mathbf{x}_0|\mathbf{z}_0) = p(\mathbf{x}_0)$ is available, the *posterior probability density* $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ is then obtained with Bayesian filters in two steps: *prediction* and *update*.

3.2 The Kalman Filter

The simplest way to solve the filtering problem is to use a *Kalman filter* [72]. Using the latter for state estimation of a dynamical system, it is tacitly assumed that the evolution and observation models are *linear*. Also, it is assumed that the noise in such models is additive and follows a Gaussian distribution with known mean and covariance. Then, the posterior density $p(\mathbf{x}_k, \mathbf{z}_{1:k})$ is also Gaussian. Hence, for linear dynamical systems, the Kalman filter provides an optimal solution to the state estimation problem with an exactly estimated posterior density.

Given these assumptions for linear systems, the *evolution* and *observation models* can be formulated as follows:

$$\text{Evolution model :} \quad \mathbf{x}_k = \mathbf{F}_k\mathbf{x}_{k-1} + \mathbf{s}_k + \boldsymbol{\epsilon}_{k-1}^{(x)} \tag{3.10}$$

$$\text{Observation model :} \quad \mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\epsilon}_k^{(m)} \tag{3.11}$$

where $\mathbf{F}$ and $\mathbf{H}$ are known matrices for the linear evolutions of the state $\mathbf{x}(t_k)$ and of the observation $\mathbf{z}(t_k)$, respectively, and $\mathbf{s}(t_k)$ is a known vector of inputs.

Assuming that the system noise $\boldsymbol{\epsilon}^{(x)}$ and measurement noise $\boldsymbol{\epsilon}^{(m)}$ have zero mean and covariance matrices $\mathbf{C}^{(\epsilon_x)}$ and $\mathbf{C}^{(\epsilon_m)}$, respectively, the *prediction* and *update* steps of the Kalman filter are given by

- **Prediction**: Responsible for projecting forward in time to obtain the *a priori* estimate at the next time step of the mean and variance.

$$\begin{aligned} \tilde{\mathbf{x}}_k &= \mathbf{F}_k\hat{\mathbf{x}}_{k-1} + \mathbf{s}_k \\ (\tilde{\mathbf{C}}_k^{(x)}) &= \mathbf{F}_k\mathbf{C}_{k-1}^{(x)}\mathbf{F}_k^T + \mathbf{C}^{(\epsilon_x)} \end{aligned}$$

- **Update**: Incorporates the new observation, $\mathbf{z}_k$, into the *a priori* estimate to obtain an improved *a posteriori* estimate.

$$\begin{aligned} \mathbf{K}_k &= (\tilde{\mathbf{C}}_k^{(x)})\mathbf{H}_k^T \left(\mathbf{H}_k(\tilde{\mathbf{C}}_k^{(x)})\mathbf{H}_k^T + \mathbf{C}^{(\epsilon_m)}\right)^{-1} \\ \hat{\mathbf{x}}_k &= \tilde{\mathbf{x}}_k + \mathbf{K}_k\left(\mathbf{z}_k - \mathbf{H}_k\tilde{\mathbf{x}}_k\right) \\ \mathbf{C}_k^{(x)} &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\left(\tilde{\mathbf{C}}_k^{(x)}\right) \end{aligned} \tag{3.12}$$

  where $\mathbf{K}$ denotes the *gain matrix* of the Kalman filter.

After estimating the state variable $\hat{\mathbf{x}}(t_k)$ and its covariance matrix $\mathbf{C}_k^{(x)}$ in the prediction step, their corresponding *a posteriori* estimates are obtained in the update step with the utilization of the measurements $\mathbf{z}(t_k)$.

3.3 Bayesian Filters

The Bayesian solution to compute the *posterior distribution*, $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, of the state vector $\mathbf{x}(t_k)$, given past observations $\mathbf{z}_{1:k}$, is given by the general *Bayesian update recursion*

$$\text{measurement update} : p\left(\mathbf{x}_k|\mathbf{z}_{1:k}\right) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{3.13}$$

$$\text{normalization} : p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int_{\mathbb{R}^N} p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})d\mathbf{x}_k \tag{3.14}$$

$$\text{time update} : p(\mathbf{x}_{k+1}|\mathbf{z}_{1:k}) = \int_{\mathbb{R}^N} p(\mathbf{x}_{k+1}|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k})d\mathbf{x}_k \tag{3.15}$$

The *posterior distribution* is the primary output from a nonlinear filter, from which standard measures as the minimum mean square estimate $\hat{\mathbf{x}}_{k|k}^{mms}$ and its covariance $\mathbf{C}_{k|k}^{(x),mms}$ can be extracted:

$$\hat{\mathbf{x}}_{k|k}^{mms} = \int \mathbf{x}_k p(\mathbf{x}_k|\mathbf{z}_{1:k})d\mathbf{x}_k \tag{3.16}$$

$$\mathbf{C}_{k|k}^{(x),mms} = \int \left(\mathbf{x}_k - \hat{\mathbf{x}}_k^{mms}\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^{mms}\right)^T p(\mathbf{x}_k|\mathbf{z}_{1:k})d\mathbf{x}_k \tag{3.17}$$

For a linear Gaussian model, the Kalman filter recursion relations provide a solution to this Bayesian filtering problem. But for non-linear or non-Gaussian models no finite-dimensional representations of the posterior density exist, and numerical approximations are needed.

3.4 Particle Filter

If linear Gaussian evolution - observation models are inadequate, the use of the Kalman filter does not result in optimal solutions. In such cases the posterior density of underlying states $\mathbf{x}_k$, given the sequence of observations $\mathbf{z}_{0:k}$ is not analytic, implying that the distribution $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ cannot be expressed in a simple form. The application of *Monte Carlo* techniques then appears as the most general and robust approach to analyze non-linear system dynamics and/or non-Gaussian state distributions. This is the case despite the availability of the so-called *extended* or *unscented Kalman filter*, which generally involves a linearization of the problem.

The *particle filter* (PF) is also known under various names as there are:

- the bootstrap filter,
- the condensation algorithm,
- interacting particle approximations,
- survival of the fittest.

The particle filter approximates the posterior distribution of unobservable states, $p(\mathbf{x}_k|\mathbf{z}_{0:k})$, with a discrete density which will be evaluated at a dynamic stochastic grid in this work.

3.4.1 Sampling on a dynamic stochastic grid

The *Particle Filter* (PF) is based on a direct application of the Bayesian recursion relations and approximates the posterior distribution, $p(\mathbf{x}_k|\mathbf{z}_{0:k})$, with a discrete density sampled on a dynamic stochastic grid. This kind of sampling has some noteworthy properties:

- A dynamic stochastic grid $\mathbf{x}_k^i$ changes over time and represents a very efficient representation of the state space.
- The PF generates and evaluates a set $\{\mathbf{x}_{1:k}^i\}_{i=1}^N$ of N different trajectories rather than current states $\mathbf{x}_k^i$ only. This affects the time update of the Bayesian recursion relations as follows:

$$\begin{aligned} p\left(\mathbf{x}_{1:k+1}^i|\mathbf{z}_{1:k}\right) &= p\left(\mathbf{x}_{1:k}^i|\mathbf{z}_{1:k}\right) p\left(\mathbf{x}_{k+1}^i|\mathbf{x}_{1:k}^i, \mathbf{z}_{1:k}\right) \\ &= w_{k|k}^i p\left(\mathbf{x}_{k+1}^i|\mathbf{x}_k^i\right) \end{aligned} \tag{3.18}$$

- The new grid is obtained by sampling from

$$p\left(\mathbf{x}_{k+1}|\mathbf{z}_{1:k}\right) = \sum_{i=1}^{N} p\left(\mathbf{x}_{1:k+1}^{i}|\mathbf{z}_{1:k}\right) = \sum_{i=1}^{N} w_{k|k}^{i} p\left(\mathbf{x}_{k+1}^{i}|\mathbf{x}_{k}^{i}\right) \tag{3.19}$$

Employing *importance sampling* amounts to introducing an auxiliary *proposal density* $q\left(\mathbf{x}_{k+1}|\mathbf{x}_{k}, \mathbf{z}_{k+1}\right)$, from which it is easy to sample. The advantage of using an auxiliary proposal distribution can be seen, if one rewrites the Bayesian recursion as follows:

$$\begin{aligned} p\left(\mathbf{x}_{k+1}|\mathbf{z}_{1:k}\right) &= \int_{\mathcal{R}^N} p\left(\mathbf{x}_{k+1}|\mathbf{x}_{k}\right) p\left(\mathbf{x}_{k}|\mathbf{z}_{1:k}\right) d\mathbf{x}_{k} \\ &= \int_{\mathcal{R}^N} q\left(\mathbf{x}_{k+1}|\mathbf{x}_{k}, \mathbf{z}_{k+1}\right) \frac{p\left(\mathbf{x}_{k+1}|\mathbf{x}_{k}\right)}{q\left(\mathbf{x}_{k+1}|\mathbf{x}_{k}, \mathbf{z}_{k+1}\right)} p\left(\mathbf{x}_{k}|\mathbf{z}_{1:k}\right) d\mathbf{x}_{k} \end{aligned} \tag{3.20}$$

Now for each particle a random sample from this auxiliary proposal distribution $\mathbf{x}_{k+1}^{i} \sim q\left(\mathbf{x}_{k+1}|\mathbf{x}_{k}^{i}, \mathbf{z}_{k+1}\right)$ is drawn, and the *posterior probability* is adjusted for each particle with the *importance weight*

$$\begin{aligned} p\left(\mathbf{x}_{1:k+1}|\mathbf{z}_{1:k}\right) &= \sum_{i=1}^{N} \frac{p\left(\mathbf{x}_{k+1}^{i}|\mathbf{x}_{k}^{i}\right)}{q\left(\mathbf{x}_{k+1}^{i}|\mathbf{x}_{k}^{i}, \mathbf{z}_{k+1}\right)} w_{k|k}^{i} \delta\left(\mathbf{x}_{1:k+1} - \mathbf{x}_{1:k+1}^{i}\right) \\ &= w_{k+1|k}^{i} \delta\left(\mathbf{x}_{1:k+1} - \mathbf{x}_{1:k+1}^{i}\right) \end{aligned} \tag{3.21}$$

Note that the proposal distribution $q\left(\mathbf{x}_{k+1}^{i}|\mathbf{x}_{k}^{i}, \mathbf{z}_{k+1}\right)$ depends on the last state in the particle trajectory $\mathbf{x}_{k}^{i}$ as well as the next measurement $\mathbf{z}_{k+1}$. The simplest choice of proposal distribution is to use the dynamic model itself, i. e. $q\left(\mathbf{x}_{k+1}^{i}|\mathbf{x}_{k}^{i}, \mathbf{z}_{k+1}\right) = p\left(\mathbf{x}_{k+1}^{i}|\mathbf{x}_{k}^{i}\right)$, leading to $w_{k+1|k}^{i} = w_{k|k}^{i}$.

Thus, the *Particle Filter* is a *sequential Monte Carlo* technique for the solution of the *state estimation problem*. It is a sampling method which approximates the *posterior distribution* by making use of its temporal structure. Again, the key idea is to represent the required *posterior density* function by a *set of random samples* (particles) with associated weights yielding a particle representation of the posterior density according to

$$p(\mathbf{x}_{k}|\mathbf{z}_{0:k}) \approx \sum_{i=1}^{N} w_{k-1}^{i} \delta\left(\mathbf{x}_{k} - \mathbf{x}_{k-1}^{i}\right)$$

where w_{k}^{i} represents the weight of particle $\mathbf{x}_{k}^{i}$. Estimates are then computed based on these samples and weights. As the number of samples becomes very large,

this Monte Carlo characterization becomes an equivalent representation of the posterior probability function, and the solution approaches the *optimal Bayesian estimate*.

Note that for the posterior density a *recursive representation* exists according to

$$p\left(\mathbf{x}_k|\mathbf{z}_{0:k}\right) = \alpha \cdot p\left(\mathbf{z}_k|\mathbf{x}_k\right) \int d\mathbf{x}_{k-1} p\left(\mathbf{x}_{k-1}|\mathbf{z}_{0:k-1}\right) p\left(\mathbf{x}_k|\mathbf{x}_{k-1}\right)$$

With a particle representation, this recursive relation simplifies to

$$p\left(\mathbf{x}_k|\mathbf{z}_{0:k}\right) \approx \alpha \cdot p\left(\mathbf{z}_k|\mathbf{x}_k\right) \sum_{i=1}^{N} w_{k-1}^i p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i\right)$$

How do we create a proper set of particles for representing the posterior distribution $p\left(\mathbf{x}_k|\mathbf{z}_{0:k}\right)$? The answer is *importance sampling* which will be explained next.

3.4.2 Sequential importance sampling

The technique of *importance sampling* is a method for generating *fair* samples of a distribution $p(\mathbf{x}(t))$. Suppose $p(\mathbf{x}(t))$ is a density from which it is difficult to draw samples, but it is easy to evaluate $p(\mathbf{x}^i(t))$ for some particular instances $\mathbf{x}^i$, i. e. on a grid. Then, an approximation to $p(\mathbf{x})$ can be given by:

$$p(\mathbf{x}) \approx \sum_{i=1}^{N} w^i \delta\left(\mathbf{x} - \mathbf{x}^i\right) \tag{3.22}$$

$$w^i = \frac{p(\mathbf{x})}{q\left(\mathbf{x}^i\right)} \tag{3.23}$$

where $q(\mathbf{x}^i)$ denotes any auxiliary *proposal distribution*, also called *importance density*. In particular, a uniform sampling of the state space could be used but would lead to *sample depletion*, or *sample degeneracy*, or *sample impoverishment*. A more direct auxiliary proposal distribution would be an approximation to the posterior density $p\left(\mathbf{x}_k|\mathbf{z}_{1:k}\right)$. This is achieved by *re-sampling*, which introduces the required *information feedback* from the observations, so trajectories that perform well, will survive the re-sampling.

A generic Particle Filter algorithm

- Choose

  - a *proposal distribution* $q\left(\mathbf{x}_{k+1}|\mathbf{x}_{1:k}, \mathbf{z}_{k+1}\right)$,
  - a *re-sampling strategy* and
  - the number of particles N.

- *Initialization*: Generate $\mathbf{x}_1^i \sim p_{x_0},\ i = 1, \ldots, N$ and set $w_{1|0}^i = \frac{1}{N}$.
- *Iteration*: for $k = 1, 2, \ldots..$ do
  - *Measurement update*: For $i = 1, 2, \ldots, N$ compute

$$w_{k|k}^i = \frac{w_{k|k-1}^i p\left(\mathbf{z}_k|\mathbf{x}_k^i\right)}{\sum_{i=1}^N w_{k|k-1}^i p\left(\mathbf{z}_k|\mathbf{x}_k^i\right)}$$

  - *Estimation*: The filtering posterior density is approximated by

$$\hat{p}\left(\mathbf{x}_{1:k}|\mathbf{z}_{1:k}\right) = \sum_{i=1}^N w_{k|k}^i \delta\left(\mathbf{x}_{1:k} - \mathbf{x}_{1:k}^i\right)$$

    and the mean is approximated by

$$\langle \mathbf{x} \rangle \approx \sum_{i=1}^N w_{k|k}^i \mathbf{x}_{1:k}^i$$

  - *Re-sampling*: Optionally at each time, take N samples with replacement from the set $\{\mathbf{x}_{1:k}^i\}_{i=1}^N$ where the probability to take sample i is $w_{k|k}^i = 1/N$
  - *Time update*: Generate predictions according to the chosen proposal density

$$\mathbf{x}_{k+1}^i \sim q\left(\mathbf{x}_{k+1}|\mathbf{x}_k^i, \mathbf{z}_{k+1}\right)$$

    and compensate for the importance weight

$$w_{k+1|k}^i = w_{k|k}^i \frac{p\left(\mathbf{x}_{k+1}^i|\mathbf{x}_k^i\right)}{q\left(\mathbf{x}_{k+1}^i|\mathbf{x}_k^i, \mathbf{z}_{k+1}\right)}$$

Note that the algorithm exhibits the fundamental structure of the Bayesian recursion relations as detailed above. Most common forms of PF algorithms combine the weight updates into one equation according to

$$w_{k|k}^i \propto w_{k-1|k-1}^i \frac{p\left(\mathbf{z}_k|\mathbf{x}_k^i\right) p\left(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i\right)}{q\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k\right)}$$

The PF algorithm outputs an approximation of the *trajectory posterior density* $p\left(\mathbf{x}_{1:k}|\mathbf{z}_{1:k}\right)$. For a filtering problem, the simplest solution would be just to extract the last state $\mathbf{x}_k^i$ from the trajectory $\mathbf{x}_k^i$ and use the *particle approximation*

$$\hat{p}\left(\mathbf{x}_k|\mathbf{z}_{1:k}\right) = \sum_{i=1}^{N} w_{k|k}^i \delta\left(\mathbf{x}_k - \mathbf{x}_k^i\right)$$

However, this is incorrect as in general all paths $\mathbf{x}_{1:k-1}^j$ can lead to the state $\mathbf{x}_k^i$. The correct solution, taking all paths leading to $\mathbf{x}_k^i$ into account, leads to an *importance weight*

$$w_{k+1|k}^i = \sum_{j=1}^{N} w_{k|k}^j \frac{p\left(\mathbf{x}_{k+1}^i|\mathbf{x}_k^j\right)}{q\left(\mathbf{x}_{k+1}^i|\mathbf{x}_k^i, \mathbf{z}_{k+1}\right)}$$

which replaces the time update of the normalized weights given in the generic PF algorithm above. This solution is called the *marginal Particle Filter* (mPF) which has many applications in system identification and robotics. Unfortunately, the complexity is now of order $O(N^2)$.

Prediction to get $p\left(\mathbf{x}_{1:k+m}|\mathbf{z}_{1:k}\right)$ can be implemented by repeating the time update in the generic PF algorithm m times.

Importance re-sampling

The basic particle filter suffers from sample depletion where all but a few particles will have negligible weights. Re-sampling solves this problem, but inevitably destroys information and thus increases uncertainty by the random sampling. In this work the *bootstrap PF*, also called *Sampling Importance Resampling* (SIR), is employed which applies re-sampling each time. To avoid sample depletion, an auxiliary *importance density* needs to be selected appropriately as the prior density $p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i\right)$.

The SIR algorithm

The SIR algorithm uses re-sampling at every iteration. It can be summarized in the following steps, as applied to the system evolution from t_{k-1} to t_k:

Step 1 – For $i = 1, \ldots, N$ draw new particles $\mathbf{x}_k^i$ from the importance density by employing the transition model

$$q(\mathbf{x}_k) = \sum_{i=1}^{N} w_{k-1}^i p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i\right)$$

To do so, choose a random number r uniformly from $[0, 1]$ and choose particle $i = r$, then sample from the prior density $p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i\right)$.

– Use the corresponding *likelihood* to calculate corresponding weights

$$w_k^i = p\left(\mathbf{z}_k|\mathbf{x}_k^i\right)$$

The samples $\mathbf{x}_k^i$, employed above, are fair samples from $p\left(\mathbf{x}_k|\mathbf{z}_{0:k-1}\right)$ and re-weighting them accounts for the evidence of the observations $\mathbf{z}_k$.

Step 2 Calculate the total weight $W_k = \sum_{i=1}^N w_k^i$.
Normalize the particle weights $w_k^i = W_k^{-1} w_k^i \ \forall i = 1, ..., N$

Step 3 Re-sample the particles by doing

– Compute the cumulative sum of weights $W_k^i = W_{k-1}^i + w_k^i; \forall i = 1, \ldots, N,\ W^0 = 0$

– Let $i = 1$ and draw a starting point u_1 from a uniform distribution $U[0, N^{-1}]$

– For $j = 1, \ldots, N$ do the following

∗ move along the cumulative sum of weights by setting $u_j = u_1 + N^{-1}(j-1)$

∗ while $u_j > W^i$ set $i = i + 1$

∗ assign samples $\mathbf{x}_k^j = \mathbf{x}_k^i$

∗ assign weights $w_k^j = N^{-1}$

The resampling procedure just described avoids to have many degenerate particles with vanishing weights but it also leads to a loss of diversity in the sense that the resulting samples may contain many redundant particles. This phenomenon is called *sample impoverishment* and is often observed in case of small process noise. In this situation, all particles collapse to a single particle within few instants t_k.

Effective number of samples

An indicator of the *degree of depletion* is the effective number of samples, defined in terms of the coefficient of variation σ_{ϵ_x} as

$$N_{eff} = \frac{N}{1 + \sigma_{\epsilon_x}^2(w_{k|k}^i)} = \frac{N}{1 + \frac{Var\left(w_{k|k}^i\right)^2}{E\{w_{k|k}^i\}}} = \frac{N}{1 + N^2 Var\left(w_{k|k}^i\right)}$$

The effective number of samples is thus at its maximum $N_{eff} = N$ when all weights are equal $w^i_{k|k} = \frac{1}{N}$, and the lowest value it can attain is $N_{eff} = 1$, which occurs when $w^i_{k|k} = 1$ with probability $1/N$ and $w^i_{k|k} = 0$ with probability $(N-1)/N$.

In practical applications this number could be approximated by

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N} \left(w^i_{k|k} \right)^2}$$

when we have $1 \leq \hat{N}_{eff} \leq N$. Again the upper bound $N_{eff} = N$ is attained when all particles have the same weight, and the lower bound $N_{eff} = 1$ when all the probability mass is devoted to a single particle. The resampling condition in the PF can now be defined as $N_{eff} < N_{th}$, and the threshold can for instance be chosen as $N_{th} = \frac{2N}{3}$.

Choice of proposal/importance distribution

The choice of proposal distribution clearly influences the depletion problem. The most general proposal distribution has the form $q\left(\mathbf{x}_{1:k}|\mathbf{z}_{1:k}\right)$. This means that the whole trajectory needs be sampled at each iteration, which in real-time applications is not realistic. But the general proposal can be factorized as

$$q\left(\mathbf{x}_{1:k}|\mathbf{z}_{1:k}\right) = q\left(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{z}_{1:k}\right) q\left(\mathbf{x}_{1:k-1}|\mathbf{z}_{1:k}\right)$$

The most common approximation in applications is to reduce the path $\mathbf{x}_{1:k-1}$ and only sample the new state $\mathbf{x}_k$, so the proposal distribution $q\left(\mathbf{x}_{1:k}|\mathbf{z}_{1:k}\right)$ is replaced by $q\left(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{z}_{1:k}\right)$, which, due to the Markov assumption, can be written as

$$q\left(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{z}_{1:k}\right) \approx q\left(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k\right).$$

The approximate proposal density predicts good values of the current state $\mathbf{x}_k$ only, not of the whole trajectory $\mathbf{x}_{1:k}$. For further insight, one needs to discuss the dependence of this proposal distribution on the SNR. Here, the SNR is defined as the ratio of the maximal value of the likelihood and prior, respectively,

$$SNR \propto \frac{max_{x_k} p(\mathbf{z}_k|\mathbf{x}_k)}{p(\mathbf{x}_k|\mathbf{x}_{k-1})}$$

Note that for a *linear Gaussian model* this yields

$$SNR \propto \sqrt{\frac{\det(\mathbf{C}^{\epsilon_x})}{\det(\mathbf{C}^{\epsilon_m})}}$$

Thus, the SNR is high if the measurement noise is small compared to the signal noise. Given this measure, one could consider different sampling strategies, but here only the one used in this work is explained.

Likelihood sampling: For *medium or high* SNR, samples are drawn from the *likelihood* $p\left(\mathbf{z}_k|\mathbf{x}_k\right)$.

The proposal distribution $q\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k\right)$ can be factorized as follows

$$\begin{aligned} q\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k\right) \approx p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k\right) &= p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i\right) \frac{p\left(\mathbf{z}_k|\mathbf{x}_{k-1}^i, \mathbf{x}_k\right)}{p\left(\mathbf{z}_k|\mathbf{x}_{k-1}^i\right)} \\ &= p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i\right) \frac{p\left(\mathbf{z}_k|\mathbf{x}_k\right)}{p\left(\mathbf{z}_k|\mathbf{x}_{k-1}^i\right)} \end{aligned} \tag{3.24}$$

Now consider the case that the likelihood $p\left(\mathbf{z}_k|\mathbf{x}_k\right)$ is much more peaky than the prior $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and if it is integrable in state space, $\mathbf{x}_k$, then one can set

$$d^2(\mathbf{x}_n, \mathbf{x}_m) p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k\right) \propto p\left(\mathbf{z}_k|\mathbf{x}_k\right)$$

Thus, a suitable proposal distribution for the *high* SNR case is based on a *scaled likelihood function*

$$q\left(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k\right) \propto p\left(\mathbf{z}_k|\mathbf{x}_k\right)$$

This choice then yields the following weight update

$$w_{k|k}^i = w_{k-1|k-1}^i p\left(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i\right)$$

Sampling from the *likelihood* requires that the likelihood function $p\left(\mathbf{z}_k|\mathbf{x}_k\right)$ is integrable with respect to $\mathbf{x}_k$. This is not the case when $N > L$, i. e. the number of unobservable states is larger than the number of observations (measurements). The interpretation in this case is that for each value of $\mathbf{z}_k$, there is an infinite-dimensional manifold of possible $\mathbf{x}_k$ to sample from, each one equally likely.

3.4.3 Resampling Schemes

The method of sampling from the *likelihood* has the drawback that it becomes unstable as k increases. There is a discrepancy between the weights, therefore the algorithm can be stabilized by performing a *re-sampling* sufficiently often. The weighted approximate density is modified by each re-sampling step. This is done by eliminating particles with low weights and multiplying particles with an important weight. The filtering posterior density then becomes:

$$\hat{p}\left(\mathbf{x}_{0:k}, \mathbf{z}_{1:k}\right) = \sum_{i=1}^{N} \frac{n_i}{N} \delta(\mathbf{x}_{0:k} - \mathbf{x}_{0:k}^i)$$

where n_i is the number of copies of particle $\mathbf{x}_{0:k}^i$. In the following, the three main *re-sampling schemes* will be described.

Multinomial Resampling

The *multinomial resampling* scheme generates N ordered uniform random numbers on the interval $(0, 1]$. The new particle set is selected according to the multinomial distribution [38].

$$\begin{aligned} u_k &= u_{k+1}\tilde{u}_k^{\frac{1}{k}},\ with\ \tilde{u}_k \sim U[0,1) \\ u_N &= \tilde{u}_N^{\frac{1}{N}} \\ \mathbf{x}_k &= x(F^{-1}(u_k)) \\ &= x_i\ with\ i\ so\ that\ u_k \in \big[\sum_{s=1}^{i-1}\omega_s, \sum_{s=1}^{i}\omega_s\big), \end{aligned}$$

where the generalized inverse of the cumulative probability distribution is F^{-1}. The weights are normalized.

Systematic resampling

With *systematic re-sampling*, the particle set is selected according to the multinomial distribution of N generated ordered numbers u_k.

$$u_k = \frac{(k-1)+\tilde{u}}{N},\ with\ \tilde{u} \sim U[0,1)$$

Residual resampling

With *residual re-sampling*, the variance can be decreased. We define:

$$N^i = \lfloor n\omega^i \rfloor + \bar{N}^i,\ \text{for } i = 1, \ldots, m,$$

where the $\bar{N}^1, ..., \bar{N}^n$ are multinomially distributed, and $\lfloor\rfloor$ means the integer part. The multinomial distribution $M(n-R; \bar{\omega}^1, ..., \bar{\omega}^n)$ has the following arguments:

$$\begin{aligned} R &= \sum_{i=1}^{m} \lfloor n\omega^i \rfloor \\ \bar{\omega}^i &= \frac{n\omega^i - \lfloor n\omega^i \rfloor}{n-R},\ i = 1, ..., m \end{aligned}$$

The multinomial counts $\bar{N}^1, ..., \bar{N}^n$ are generated in the same way as in the *multinomial re-sampling* described above.

Chapter 4

Bi-dimensional EMD

In nuclear medicine radiation therapy presupposes an estimation of an absorbed energy dose distribution from known X-ray CT-based mass density distributions combined with SPECT-PET-based radioactivity distributions. Such mappings can be learned employing *Deep Neural Networks* (DNN). However, such deep neural architectures are complex and training such networks is computationally expensive. The current study elaborates on exploratory image analysis techniques which try to extract characteristic features from given mass density distributions. Such features might represent a simpler yet informative input to a DNN and may reduce the computational load considerably.

Traditionally, exploratory feature extraction techniques like *Principal Component Analysis* (PCA) [24, 28] and Non-negative Matrix and Tensor Factorization [29] offer a global rather than local analysis. However, these methods are built upon assumptions and constraints (orthogonality, statistical independence) which might not always be obeyed. Generally, spatial variations of mass density distributions as well as spatio-temporal variations of activity distributions in fMRI images represent non-linear and non-stationary signals while the techniques mentioned require at least wide-sense stationary signals. A decade ago, [62] proposed an empirical method to analyze such data sets called *Empirical Mode Decomposition* (EMD). Together with a Hilbert transform, this method allows to study the instantaneous frequency content of signals. Its noise - assisted variant, called EEMD represents a powerful tool for analyzing biomedical time series.

Concerning biomedical images, obviously, two-dimensional data arrays were of special interest [126]. With *Bi-Dimensional EMD* (BEMD), it is generally required to find local maxima and minima and subsequently interpolating these points in each iteration of the process [88, 110, 31, 91, 167, 109]. In [31] interpolation methods based on a Delaunay triangulation with subsequent cubic interpolation on triangles has been studied. Instead of considering envelope surface interpolation, Bhuiyan et al. [15, 16] studied both a direct envelope surface es-

timation method and radial basis function interpolators. Similarly, Xiong et al. [167] implemented a FastRBF algorithm for the estimation of the envelope surfaces. However, detection and interpolation of local extrema render the process complicated and time consuming. A major breakthrough has been achieved by Wu et al. [166], who proposed a *Multi-dimensional Ensemble Empirical Mode Decomposition* (MEEMD) for multidimensional data arrays. MEEMD turned out to be the best choice in practical applications, for example, medical image analysis [6], image analysis [108], texture analysis [92], laser speckle contrast images [65] etc..

In this study, a novel and computationally highly efficient BEMD approach is considered, which replaces the direct cubic spline envelope surface interpolation step by an interpolation based on Green's function with tension [164, 5]. This technique is applied in this study to images representing mass density distributions for extracting *Bi-dimensional Intrinsic Mode Function* (BIMF)s on various spatial scales. These characteristic textures might be usefull on the one hand to estimate the number of filters to be applied in the *Convolutional Neural Network* (CNN)s or to use them immediately as input to the DNN.

4.1 Canonical Bi-dimensional EMD

EEMD tries to extract from any multi-variate signal simple oscillations locally. These component signals are called *Intrinsic Mode Function* (IMF), henceforth called intrinsic modes for short. In case of images such spatial oscillations can be seen as characteristic textures of the given intensity distribution. Note that intensities can only be positive, hence to apply EMD, images need to be centered beforehand. IMFs are obtained from the signal through a process called sifting, which results in pure oscillations with zero-mean but time-varying amplitude and frequency. Furthermore, IMFs are ordered according to their frequency content. In contrary to wavelet analysis, EMD is a data driven algorithm that decomposes the signal without prior knowledge.

The decomposition of an image starts by vectorizing it into a row vector $\tilde{\mathbf{x}}_m \in \mathbb{R}^N$. All images are then collected row-by-row into a data matrix $\mathbf{X}^{M\times N}$. Next EEMD is applied to each column $X_{*n} \equiv \mathbf{x}_n$ of the $M \times N$ - dimensional data matrix $\mathbf{X}$, where M denotes the number of samples and N gives the dimension of the data vectors, i. e. the number of pixels in every image. The 1D-EEMD of the n-th column becomes

$$\mathbf{x}_n := X_{*,n} = \sum_{j=1}^{J} C_{*,n}^{(j)} \tag{4.1}$$

where the column vector $C_{*,n}^{(J)}$ represents the residuum of the n-th column vector of

the data matrix. This finally results in J component matrices, each one containing the j-th component of every column $\mathbf{x}_n, n = 1, \ldots, N$ of the data matrix $\mathbf{X}$.

$$\mathbf{C}^{(j)} = [\mathbf{c}_1^{(j)}\ \mathbf{c}_2^{(j)}\ \cdots\ \mathbf{c}_N^{(j)}] = [C_{*,1}^{(j)}\ C_{*,2}^{(j)}\ \cdots\ C_{*,N}^{(j)}] \tag{4.2}$$

Next one applies an EEMD to each row of eqn. 4.2 yielding

$$C_{m,*}^{(j)} = \left(c_{m,1}^{(j)} c_{m,2}^{(j)} \cdots c_{m,N}^{(j)}\right) = \sum_{k=1}^{K} \left(h_{m,1}^{(j,k)} h_{m,2}^{(j,k)} \cdots h_{m,N}^{(j,k)}\right) = \sum_{k=1}^{K} H_{m,*}^{(j,k)} \tag{4.3}$$

where $c_{m,n}^{(j)} = \sum_{k=1}^{K} h_{m,n}^{(j,k)}$ represents the decomposition of the rows of matrix $\mathbf{C}^{(j)}$. These components $h_{m,n}^{(j,k)}$ can be arranged into a matrix $\mathbf{H}^{(j,k)}$ according to

$$\mathbf{H}^{(j,k)} = \begin{bmatrix} h_{1,1}^{(j,k)} & h_{1,2}^{(j,k)} & \cdots & h_{1,N}^{(j,k)} \\ h_{2,1}^{(j,k)} & h_{2,2}^{(j,k)} & \cdots & h_{2,N}^{(j,k)} \\ \vdots & \vdots & \cdots & \vdots \\ h_{M,1}^{(j,k)} & h_{M,2}^{(j,k)} & \cdots & h_{M,N}^{(j,k)} \end{bmatrix} \tag{4.4}$$

The resulting component matrices have to be summed to obtain

$$\mathbf{C}^{(j)} = \sum_{k=1}^{K} \mathbf{H}^{(j,k)}. \tag{4.5}$$

Finally this yields the following decomposition of the original data matrix $\mathbf{X}$

$$\mathbf{X} = \sum_{j=1}^{J} \mathbf{C}^{(j)} = \sum_{j=1}^{J} \sum_{k=1}^{K} \mathbf{H}^{(j,k)} \tag{4.6}$$

where each element is given by

$$x_{m,n} = \sum_{j=1}^{J} \sum_{k=1}^{K} h_{m,n}^{(j,k)} \tag{4.7}$$

To yield meaningful results, components $h_{m,n}^{(j,k)}$ with comparable scales, i.e. similar spatial frequencies of their textures, should finally be combined [166] according to a *Comparable Minimal Scale Combination Principle* (CMSC). In practice, for two-dimensional data sets this implies that the components of each row, which represent a common horizontal scale, and the components of each column, which represent a common vertical scale, should be summed up [166].

Hence, the CMSC - principle leads to BIMFs given by

$$\mathbf{S}^{(k')} = \sum_{k=1}^{K} \mathbf{H}^{(k,k')} + \sum_{j=k'+1}^{J} \mathbf{H}^{(k',j)} \tag{4.8}$$

which thus yields a decomposition of the original data matrix $\mathbf{X}$ into BIMFs according to

$$\mathbf{X} = \sum_{k'=1}^{K} \mathbf{S}^{(k')} \tag{4.9}$$

where $\mathbf{S}^{(K)}$ represents the non-oscillating residuum. The extracted BIMFs can be considered features of the data set which, according to the CMSC - principle, reveal local textures with characteristic spatial frequencies which help the further analysis. The latter can be improved using a noise-assisted variant, which mainly helps to avoid mode mixing due to local frequency degeneration.

4.2 Noise-assisted BEMD

Each CT slice is encoded as a two-dimensional array of intensity values $\mathbf{X} = \{x_{mn}\}, m = 1, \ldots, M; n = 1, \ldots, N$. These pixels represent discrete intensity values at the coordinates $x_m \equiv m, y_n \equiv n$ of a continuous intensity distribution $f(x, y)$. Before image decomposition, pixel values are z-scored, i. e. normalized to zero-mean and unit variance.

Next a special variant of a noise - assisted EEMD, starting with two noisy versions of every image, is proposed. First white noise with zero mean and standard deviation $\sigma_n = 0.2$ is added and subtracted, thus forming the two versions $\tilde{f}^*(x, y)$: $\tilde{f}^+(x, y) = f(x, y) + \varepsilon$ and $\tilde{f}^-(x, y) = f(x, y) - \varepsilon$. Therefore, the EMD of each noisy input image yields

$$\tilde{f}^*(m, n) = \sum_{j=1}^{J} c_j^*(m, n) \tag{4.10}$$

where $c_J^*(m, n)$ denotes the residuum, and the $c_i^*(m, n)$ designate the noisy bi-dimensional intrinsic modes. Note that by averaging the two noisy versions $0.5(f^+(x, y) + f^-(x, y))$, the original image $f(x, y)$ is obtained. Therefore performing the BEEMD with a fixed number J of *Bi-dimensional Intrinsic Mode* (BIM), each BIM of $f(x, y)$ is naturally obtained by averaging the corresponding BIMs of the noisy versions. The main steps of this noise-assisted BEMD [108, 102], [5] are as follows:

1. Assign the number of intrinsic modes (BIMs) $J = 6$
   Assign the number of sifting steps $L = 5$
   Initialize $c_1^*(m, n) = \tilde{f}(m, n)$.

2. Extract the $j = 0, 2 \ldots J$ BIMs

   (a) Do the sifting by repeating the following steps:

i. Extract all local 2D-maxima and 2D-minima of $\hat{c}^*(x, y)$.

ii. Build a 2D surface interpolation by employing the extrema from the previous step to obtain a 2-D envelope surface $s_{max}(x, y)$ of maxima, and a 2-D envelope surface $s_{min}(x, y)$ of minima. Calculate the average of the envelope surfaces $\langle s(x, y)\rangle$

$$\langle s(x, y)\rangle = 0.5(s_{max}(x, y) + s_{min}(x, y))$$

iii. Update function $\hat{c}^*(x, y)$ by subtracting the average surface $\langle s(x, y)\rangle$ from the $\hat{c}^*(x, y)$;

$$\hat{c}^*(x, y) \leftarrow \hat{c}^*(x, y) - \langle s(x, y)\rangle$$

(b) If the sifting loop doesn't finish, go to step 2(a)i. Otherwise a BIM is

$$c_i^*(x, y) = \hat{c}^*(x, y)$$

3. Calculate the new $\hat{c}^*(x, y)$ by subtracting all the computed BIMs from the noise - enhanced image.

$$\hat{c}^*(x, y) = \tilde{f}^*(x, y) - \sum_{i<(i+1)} c_i^*(x, y)$$

4. If $(i + 1) < I$, go to step 2 and estimate the next $(i + 1)$ BIM, otherwise $\hat{c}^*(x, y) = c_I^*(x, y)$ is the residuum.

Common ensemble decomposition algorithms apply this procedure to an ensemble of E noisy versions of the original image with concomitant increase in computational costs.

Employing *Green's function in Tension - BEEMD* (GiT-BEMD), only two noisy versions of each image are needed, reducing the computational load dramatically.

4.3 Green's function-based BEEMD

A major drawback of this or similar BEMD techniques is that the number of BIMs and their characteristics are highly dependent on the envelope estimation techniques used in the sifting process, on the methods to detect extrema, and on stopping criteria during the iterations. The following section presents a new variant of BEMD, which is based on a Green's function expansion of the interpolating

surfaces including surface tension and containing all extremal values of the bi-dimensional data set. Introducing a tension parameter alleviates surface interpolation problems and greatly improves the stability of the method relative to gridding without tension. GiT-BEMD thus differs from the canonical BEEMD algorithm basically in the process of robustly estimating the upper and lower surfaces, and in limiting the number of iterations per BIMF to a few iterations only. The details of the extrema detection and surface formation of the GiT-BEMD process are discussed in the following section according to [5].

4.3.1 Extraction of local extrema

Local extrema are points that have the largest or smallest pixel values relative to their K-connected neighbors, therefore in a $2D$ image the pixel (m, n) with coordinates (x_m, y_n) has $K = 8$ connected neighbors with coordinates $(m \pm 1, n), (m, n \pm 1), (m \pm 1, n \pm 1), (m \pm 1, n \pm 1)$. The $2D$ array of local maxima is called a *maxima map*, and the $2D$ array of local minima is called a *minima map*, respectively. A neighboring window method is employed to detect local extrema during intermediate steps of the sifting process for estimating a BIM of any source image. In this method, a data point/pixel is considered a local maximum (minimum), if its value is strictly higher (lower) than all of its neighbors. Let $P = \{P_i | i = 1, ...I\}$ be a set of local minima (maxima) of an $M \times N$- dimensional data matrix $\mathbf{X}$ such that it exists a small (large) neighborhood around any such local optimal point P_i on which the pixel value is never larger (smaller) than $f(x_i, y_i)$ at P_i. Local extrema occur only at critical points. Let $D(m, n) = f_{mm} f_{nn} - (f_{mn})^2$. If $D > 0$ at a critical point, then the critical point P_i is a local extremum. The signs of f_{mm} and f_{nn} determine whether the point is a maximum or a minimum. If $D \leq 0$ at a critical point, then the point P_i is a saddle point. Though, in practice, a 3×3 window results in an optimal *extrema map* for a given $2D$ image for many applications, sometimes a larger window size is suitable especially with large scale images.

4.4 Green's function for estimating envelopes

In BEEMD, spline interpolation is basically used to find the smoothest surface passing through a grid of irregularly spaced extrema, either maxima or minima. With GiT-BEMD, Green's functions are employed instead. The latter are deduced from proper data constraints to expand the interpolating surface under tension. Thus the envelope surfaces connecting local extrema in 2D space are determined to minimize the curvature of the surface in the presence of surface tension [164].

Interpolation with Green's functions implies that the points of the interpolating envelope surface can be expressed as

$$s(\mathbf{x}_u) = \sum_{v=1}^{V} w_v \Phi(\mathbf{x}_u, \mathbf{x}_v) \tag{4.11}$$

where $V = MN$, $\mathbf{x}_u$ denotes any point where the surface is unknown, $\mathbf{x}_v$ represents the v-th data constraint, $\Phi(\mathbf{x}_u, \mathbf{x}_v)$ is the Green's function and w_v is the respective weight in the envelope representation. Several works discuss the use of Green's functions in interpolation problems (see for example [163]). Following, the basics of the Green's function method for spline interpolation using surface tension is summarized and closely follows [5].

The Green's function $\Phi(\mathbf{x})$ obeys the following relation at any data constraint $\mathbf{x}_v,\ v = 1, \ldots, V$ [134]

$$\left[D\Delta_{op}^2 - T\Delta_{op}\right] \Phi(\mathbf{x}_u, \mathbf{x}_v) = \delta(\mathbf{x}_u - \mathbf{x}_v) \tag{4.12}$$

where, Δ_{op}^2 and $\Delta_{op} = \nabla^2$, denote the bi-harmonic, the Laplace and the Nabla operator, respectively, D is the flexural rigidity of the curve or surface, T is the tension used at the boundaries, and $\Phi(\mathbf{x}_u, \mathbf{x}_v)$ represents the Green's function containing the spatial position vectors $\mathbf{x}_u, \mathbf{x}_v$ as argument. With vanishing surface-tension, i. e. $T \to 0$, the minimum curvature solution $\Phi(\mathbf{x}_u, \mathbf{x}_v) = x_{uv}^2 \log(x_{uv}),\ x_{uv} = |\mathbf{x}_u - \mathbf{x}_v|$ is achieved [134]; while in case of a vanishing surface rigidity, i. e. $D \to 0$, the solution approaches $\Phi(\mathbf{x}_u, \mathbf{x}_v) = \log(x_{uv})$. The general solution is expected to retain these limiting characteristics. To obtain the former, rewrite eqn. 4.12 in terms of the curvature $\Psi(\mathbf{x}) = \nabla^2\Phi(\mathbf{x})$ of the Green's function and transform it to the conjugate Fourier domain where it then reads

$$\left[\Delta_{op} + \frac{p^2}{k^2}\right] \Psi(\mathbf{k}) = -\frac{1}{T}\frac{p^2}{k^2}. \tag{4.13}$$

Here $k = |\mathbf{k}|$ represents the radial wave number, $p^2 = \frac{T}{D}$, $\mathbf{k}$ denotes the wave number vector and $\Psi(\mathbf{k})$ represents the Fourier transform of $\Psi(\mathbf{x})$. In Fourier space, the solution is obtained as

$$\Psi(\mathbf{k}) = -\frac{1}{T}\left(\frac{p^2}{k^2 + p^2}\right) \tag{4.14}$$

From this the general solution of eqn. 4.12 in a 2-D spacial domain [164] can be achieved by using the inverse Hankel transform as

$$\Psi(\mathbf{x}) = -\frac{1}{T}\int_0^\infty \frac{p^2}{k^2 + p^2} J_0(kx) k dk = -\frac{1}{T} p^2 K_0(px) \tag{4.15}$$

where K_0 denotes the modified Bessel function of the second kind and order zero given by

$$K_0(px) = \int_0^\infty \frac{\cos(kpx)dk}{\sqrt{k^2+p^2}} \propto \begin{cases} \exp(-px) & \text{if} \quad px \to \infty \\ -\log(px) & \text{if} \quad px \to 0 \end{cases} \tag{4.16}$$

Integrating $\Psi(\mathbf{x})$ twice and rescaling, finally, yields the Green's function $\Phi(\mathbf{x})$ and its local gradient $\nabla\Phi(\mathbf{x})$ as

$$\begin{aligned} \Phi(\mathbf{x}_u, \mathbf{x}_v) &= log(p|\mathbf{x}_u - \mathbf{x}_v|) + K_0(p|\mathbf{x}_u - \mathbf{x}_v|) = \log(px_{uv}) + K_0(px_{uv}) \quad (4.17) \\ \nabla\Phi(\mathbf{x}_u, \mathbf{x}_v) &= p \cdot \left[\frac{1}{px_{uv}} - K_1(px_{uv})\right] \cdot \frac{(\mathbf{x}_u - \mathbf{x}_v)}{|\mathbf{x}_u - \mathbf{x}_v|} = p \cdot \left[\frac{1}{px_{uv}} - K_1(px_{uv})\right] \cdot \mathbf{e}_{uv} \end{aligned}$$

where $p \propto T$ represents the tension parameter, $|....|$ denotes the Euclidean distance, $K_0(.)$ represents the modified Bessel function and $\mathbf{e}_{uv}$ denotes the unit vector pointing along the direction $\mathbf{x}_u - \mathbf{x}_v$. Hence, by decreasing the tension parameter $p \propto T$, the solution is expected to reach the minimum curvature solution represented by the bi-harmonic Green's function [134]. In contrary, increasing the tension parameter T, thus also p, renders the arguments of $\Phi(\mathbf{x})$ large and leads to an interpolating surface dominated by tension. Thus varying the tension p achieves a continuous spectrum of Green's functions reflecting the trade-off between the minimum curvature solution driven by the $\log(px)$ term and the impact of the surface tension via the modified Bessel function $K_0(px)$. Finding an optimal tension parameter is still an open problem. If intrinsic data modes are to be used for classification purposes, an optimal tension parameter could be identified as the one achieving maximal classification accuracy. The effect of the tension parameter is illustrated in Fig. 4.1.

Including V data constraints yields for the defining equation and its solution [134]

$$\begin{aligned} \Delta_{op}\left[\Delta_{op} - p^2\right] c(\mathbf{x}_u) &= \sum_{v=1}^{V} w_v \delta(\mathbf{x}_u - \mathbf{x}_v) \\ c(\mathbf{x}_u) &= \sum_{v=1}^{V} w_v \Phi(\mathbf{x}_u - \mathbf{x}_v) \end{aligned} \tag{4.18}$$

The coefficients w_v can be obtained by solving the system of linear equations $\mathbf{Gw} = \mathbf{c}$ where the Green's matrix $\mathbf{G}$ collects all Green's functions $\Phi(\mathbf{x}_m - \mathbf{x}_v)$ at the data constraints $v = 1, \ldots, V$. Corresponding slopes s_m in directions $\hat{\mathbf{n}}_m$ can be obtained by evaluating the relations

$$s_m = \sum_{m=1}^{N} w_m \nabla\Phi(\mathbf{x}_m - \mathbf{x}_n) \cdot \hat{\mathbf{n}}_m \qquad m = 1, \ldots, N.$$

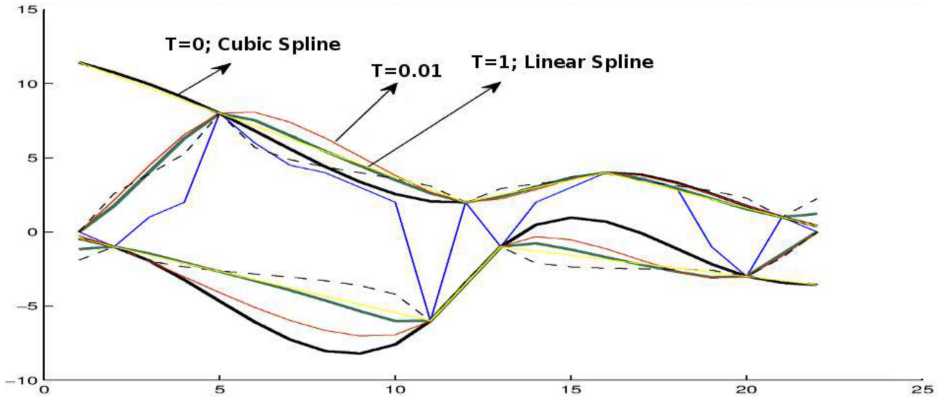


Figure 4.1: The figure shows the three splines along with their tension values that pass through the same extremal (local maxima and minima) points. Note that when the tension vanishes $T = 0$, the points are connected by a cubic spline; and when $T = 1$, it is representing a linear spline.

In summary, the interpolation procedure is based on two steps: the first step estimates the weights $\mathbf{w} = [w_1 \quad w_2 \dots w_P]$ and the second step estimates the interpolating envelope surface:

- The surface values $\mathbf{s}(\mathbf{x}_v) = [s(\mathbf{x}_1), \dots, s(\mathbf{x}_V)]^T \equiv \mathbf{c} = [c_1, c_2, \dots, c_V]^T$ are known in a total of $V = MN$ locations $\mathbf{x}_v = [x_m, y_n]^T$. Then using the interpolation equation 4.11 for each of the known points, a linear system with V equations is obtained

$$\mathbf{G}\mathbf{w} = \mathbf{c}$$

  where the m-th row of matrix $\mathbf{G} \in \mathbb{R}^{M \times N}$ is the evaluation of the Green's function $\Phi(\mathbf{x}_u, \mathbf{x}_v), v = 1, 2 \dots V$. The corresponding weights are then obtained as $\mathbf{w} = \mathbf{G}^{-1}\mathbf{c}$.

- Using the weights $\mathbf{w}$, the value $s(\mathbf{x}_u) \equiv c_u$ of the envelope surface can be estimated at any point $\mathbf{x}_u$ by solving equation 4.11, which can be re-written as

$$c_u = \mathbf{w}^T \mathbf{\Phi} \tag{4.19}$$

  where the vector $\mathbf{\Phi} = [\Phi(\mathbf{x}_u, \mathbf{x}_1) \quad \Phi(\mathbf{x}_u, \mathbf{x}_2) \dots \Phi(\mathbf{x}_u, \mathbf{x}_V)]^T$ contains the Green's function values of all distances between the V data constraints and the considered location.

4.5 Dimension reduction by PCA

Having achieved a decomposition of the mass density distribution into its underlying BIMs, the latter represent characteristic textures, which are confined to specific spatial frequency bands and are nearly orthogonal locally but not globally. To achieve global orthogonality, a PCA may be performed on the matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ containing all intrinsic modes of the mass density distribution. The goal of PCA is to find a smaller, yet informative set of variables with less redundancy which represents the original signal as accurately as possible [66]. PCA performs an eigendecomposition of the centered data correlation matrix $\mathbf{Cor} = \mathbf{X}^T\mathbf{X}$ with orthogonal eigenvectors which span the input space and ordered eigenvalues which represent corresponding data variances along these new coordinates. By deliberately neglecting eigenvectors belonging to small eigenvalues during reconstruction, a more compact representation of the data can be achieved with minimal information loss in a least mean squares sense. Thus PCA can be employed to reduce the dimensionality of the mass density textures extracted by GiT-BEMD.

Chapter 5

Neural networks (NNs)

Artificial *Neural Network* (NN) represent graphical models which form the substrate for flexible learning systems. Depending on their structure they can implement different learning modalities like supervised or unsupervised learning, correlation-based or competitive learning, reinforcement learning, imitation and transfer learning etc. Feed-forward networks resemble directed acyclic graphs with nodes and edges and one or more hidden layers of nodes which have no direct contact to the outside world. Accordingly they correspond to shallow [57], [95] or deep architectures [172], [49], [2]. While the former are generally fully connected, the latter are not. Rather they implement the concept of receptive fields, which restrict the region in input space from where any neuron can receive stimuli. During repeated stimulus presentation, the weights of the network become trained in a supervised fashion if a set of labelled training data is available. Given any appropriate objective function at the output, weight adaptation is performed through gradient descent learning, which seeks to minimize the given objective function.

5.1 Gradient descent optimization in NNs

5.1.1 General

First of all a few general statements concerning *gradient descent learning* in neural networks [57], [130]:

- Gradient descent is an optimization algorithm for finding the weights of machine learning algorithms such as NNs.
- It works by having the model making predictions on training data and using the error on the predictions to update the model in such a way as to reduce the error. This is called *error-backpropagation*.

- The training error of a NN as a function of its weights is highly *non-convex*.
- The problem of training NNs is *NP-hard*
- With a finite number of functions, say N, every training error will be close, by a small constant factor, to every test error once you have more than $\log N$ training cases.
- If every training error is close to its test error, then overfitting is basically impossible.

5.1.2 Gradient descent

Gradient descent learning comes in three variants [49]:

- *Incremental (stochastic) learning*
  - Present one randomly chosen training example and then update the parameters according to the prediction error, then choose the next training sample randomly and so on. If one has presented all training samples once, an epoch is finished.
  - Stochastic learning can escape from local optima, is often fast and shows learning performance immediately. But stochastic learning is noisy so it is harder to settle in a local optimum, it may take longer on large datasets and can have a high variance over the training set.
- *Minibatch learning*
  - One partitions its training dataset (the batch) into subgroups (the minibatches), presents one minibatch and then updates the parameters according to the total error accumulated with the minibatch, then one repeats this procedure for every minibatch. When all minibatches were presented, an epoch is finished.
  - Minibatch learning can avoid local optima while having more stable gradients and is computationally efficient as a good compromise between incremental and batch learning. But minibatch learning needs an additional hyperparameter (the minibatch size) and needs part of the training dataset to be stored in memory. A good default minibatch size seems to be 32 according to literature reports. Minibatch size can be optimized independent from all other hyperparameters. Minibatch size and learning rate should be optimized after all other hyperparameters and then batch size should be fixed while learning rate can still be adapted further.

- *Batch learning*
  - One presents all training data once as a batch and then updates the parameters according to the total error of the prediction. This is called an epoch.
  - Batch learning may be more efficient due to lesser updates, has more stable gradients and is easy to parallelize. But batch learning may experience premature convergence, it may get stuck in a local optimum, the entire training dataset needs to be available in memory during learning and training becomes slow on large datasets

5.1.3 RMSProp

Stochastic Gradient Descent (SGD) is a stochastic approximation to the gradient descent optimization [59]. It is an iterative method to optimize a *differentiable* objective function. Hence the task is as follows: Minimize the following objective function

$$Q(w) = \frac{1}{N}\sum_{n=1}^{N} Q_n(w)$$

Each summand is typically associated with the n-th observation in the training data set. Such sum-minimization problems arise, for example, in least squares minimization problems, *Maximum Likelihood Estimation* (MLE) or empirical risk minimization. Minimizers of sums are called M-estimators.

- A *batch gradient descent method* would iterate the following

$$w = w - \eta \nabla_w Q(w) = w - \frac{\eta}{N}\sum_{n=1}^{N} \nabla_w Q_n(w)$$

  with η the step size or learning rate.

- A *minibatch gradient descent method* would iterate the following

$$w = w - \eta \nabla_w Q(w) = w - \frac{\eta}{K}\sum_{n=1}^{K} \nabla_w Q_n(w),\ K \ll N$$

  with η the variable step size or learning rate (see next item).

- A *stochastic gradient descent method*, instead, approximates the above by

$$w = w - \eta \nabla_w Q_n(w)$$

  where an adaptive learning rate is generally used.

- With RMSProp [59], [130] the learning rate for each weight is divided by a running average of the magnitudes of recent gradients for that weight
  - compute running average

$$v(w,t) = \gamma v(w,t-1) + (1-\gamma)\left(\nabla_w Q_n(w)\right)^2$$

    where γ is a forgetting factor
  - weight parameter update

$$w = w - \frac{\eta}{\sqrt{v(w,t)}} \nabla_w Q_n(w)$$

5.2 Deep Learning

The flexibility of neural networks is a very powerful property [138], [54], [86]. DNN profit from a *multi-stage processing*. In addition,

- DNNs possess the ability to handle nonlinear data
- DNNs allow for many modifications.

The major difference of DNN compared to shallow nets, for example *Multi-Layer Perceptrons* (MLP), is the use of the concept of receptive fields. The latter denoted the region in input space from where every node in a given processing layer of a NN receives input stimuli. Whereas in shallow feedforward neural networks all nodes in a subsequent layer are fully connected to all nodes in the preceding layer, in a DNN each node of a subsequent layer only receives input from a limited region of the preceding layer. Information processing then proceeds via a convolution operation.

5.2.1 2D Convolutions

The operation

The convolution operation uses a 2D data array, i. e. a matrix, which is often called a *kernel* or a *filter* [49]. The matrix elements represent weights. The kernel

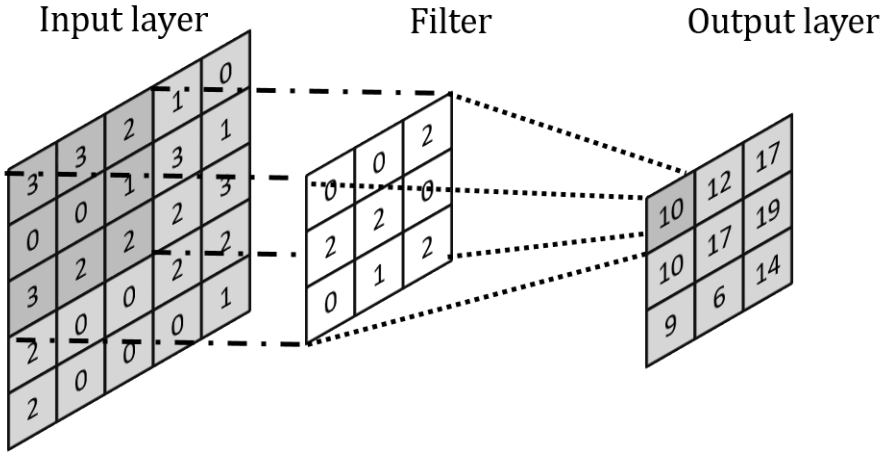


Figure 5.1: Convolutive filtering

slides over the 2D input data array, often an image, and multiplies elementwise the entries of the input data array with the corresponding weights of the kernel (see figure 5.1). Finally a weighted sum is computed which represents the output of the filter, i. e, the result of the convolution operation. If the stride of the kernel is larger than one, dimension reduction is achieved as well.

Convolutive filtering thus combines information locally, often called context, available in the preceding layer (input) into a new feature represented in the node at the subsequent layer (output). The feature, which is extracted thereby, obviously depends on the weights of the filter as well as its size. In that way the sliding kernel generates a new feature map, where each node also encodes the local context of any node of the preceding layer. This local information processing drastically reduces the number of kernel weights (or filter coefficients) to be adapted or learnt. Note that input location and output location roughly correspond to each other, keeping locality largely preserved.

To illustrate the convolution operation in 3D, imagine we start with a $9 \times 9 \times 9$ input array and use $N_F = 4$ filters with an RF size $s = 3 \times 3 \times 3$. If we also include padding with one sheet of voxels around the data cube, we get $N_{FM} = 4$ new feature maps with equal size i. e. encompassing 9^3 voxels, after the first convolution step. If we repeat filtering, using different filters, of course, we would get $N_{FM} = N_F \times N_{FM} = 16$ new feature maps. If, instead, we would only use one type of filter for this second convolution layer, hence we would get $N_{FM} = 4$ new feature maps. Even more restrictive, we could use for the second convolution step the same filters as used in the first convolution step. We then would focus on the same features, but on an increasing scale compared to the preceeding layer,

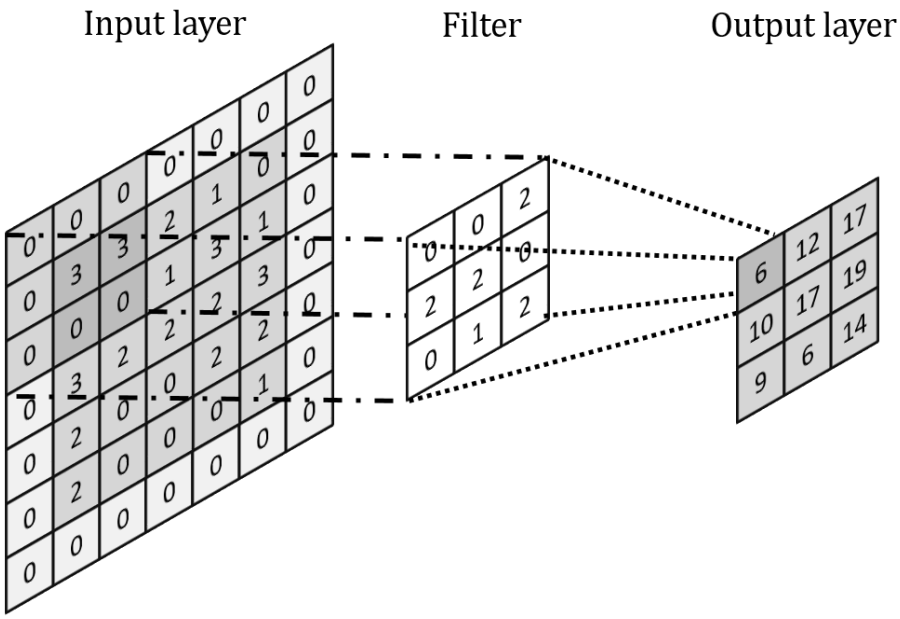


Figure 5.2: Zero padding

i. e. the receptive field size would be increased.

Padding, pooling and strides

Keeping the output feature map the same size as the input feature map is only possible, if at the borders of the input array an operation called *padding* is performed, whereby extra pixels (or voxels) are added along the borders, depending on the filter size. If the filter represents an array of 3×3 pixels, one row of extra pixels needs to be added. If, instead, the filter array has size 5×5, two rows of extra pixels would be needed (see figure 5.2). For *zero-padding*, the activities of these extra pixels are commonly chosen to be zero. These extra pixels allow every border pixel to be the center of the kernel and to be mapped to the subsequent feature map.

If, however, during filtering the spatial dimension should to be reduced, this could be achieved conveniently with a *pooling* operation. For example, a 2×2 region of the input array could be pooled to one output pixel by either computing the mean of the pixels in this region or by simply mapping the largest value of it for the output pixel (average or max pooling).

Spatial contraction could be achieved also via *striding*, which means skipping some of the slide locations of the kernel. For example, a stride of 2 only picks

every second slide and skips the one between. This downsizes the output roughly by a factor of two. A stride of 3 skipps two slides between the ones to be picked, downsizing by a factor of three roughly etc. Recent deep learning architectures favor strides over pooling, because the convolution operation still keeps some of the "skipped" information.

Multi-channel versions

Given the application we consider in this study, we choose to have two input channels: one providing the mass density maps and the other providing the activity distributions from the SPECT or PET images. Thus a filter represents a collection of two kernels, where each channel has its own unique kernel, and each such kernel produces one and only one layer of new feature maps. During convolution, each of the kernels of the filter slides over its respective input channel to produce a filtered version of it. The weights of the different kernels are in general different, thus allowing to put more emphasis on one input channel than on the others. The filter as a whole finally produces one overall output channel by summing up all the processed versions. In addition, each output filter has one bias term, which gets added to the output channel.

To summarize, each filter processes its input with its own, different set of kernels and a scalar bias term to produce a single output channel. The latter are then concatenated together to produce the overall output. The number of output channels thus equals the number of filters. A non-linear transformation, i. e. an activation function, is subsequently applied, before the output is input to the next convolutional layer for further processing.

Locality

Kernels have to learn weights from only a set of local inputs. Hence striding the kernel across all the input image, it will learn features general enough to be found at any location inside the image. Consider an edge detector like a sobel filter (see Fig. 5.3). It receives only local input, but is applied across an entire image and everywhere detects the same feature, a contrast edge. But a small kernel cannot detect an entire face, for example. Now remember that convolved images are still images. Hence, convolution operations, using filter kernels, can be stacked. It is here, where the concept of a receptive field comes in.

Receptive fields

The concept of receptive fields determines the context which is locally represented in any node of the subsequent feature map. Through striding the filter kernel

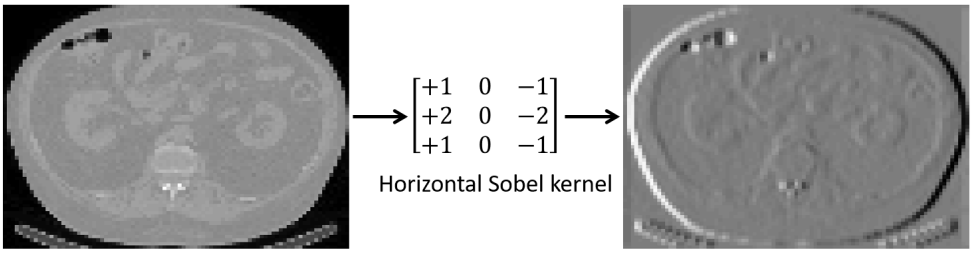


Figure 5.3: Sobel edge detector

across the entire input feature map, its weights get trained to detect always the same feature in all the input array. On the other hand striding or pooling reduces the size of the subsequent feature map whereby the receptive field gets larger and larger although the size of the kernel stays the same in all convolution layers. Thus each single pixel in the output is a representative of a larger and larger area from the same rough location from the original input. In that way, the whole convolutional filtering process increases the number of filters but each subsequent filter is characterized by larger and larger receptive fields until the latter extends over all the input image.

In summary, by detecting low level features first and progressively combine them to higher level features through convolutions and striding, it eventually detects entire visual concepts such as faces, trees, birds or other objects.

5.2.2 ReLUs in DNNs

While error back-propagation learning in shallow NNs employ sigmoidal activation functions for their hidden layer nodes, activations of *hidden* neurons in DNNs need not be bounded. Hence *Leaky Rectified Linear Units* (ReLU), $g(\mathbf{z}) = \max(\alpha, \mathbf{z})$ can replace sigmoidal activation functions like the logistic function (physicists call it Fermi function). ReLUs [105], [1] allow for better approximation quality as they avoid the *flat spot problem* during Backprop learning:

- Note that gradient estimation in layer l amounts to multiply the local gradient in layer $l + 1$ with the partial derivative of the sigmoidal activation function.
- The latter has small values except when the input is close zero.
- Small gradients result in inefficient learning.

To see this, compare the gradient of a logistic vs a ReLU activation function in figure 5.4.

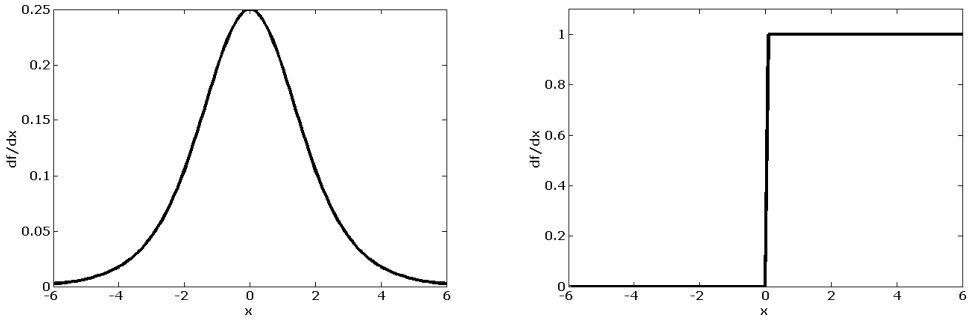


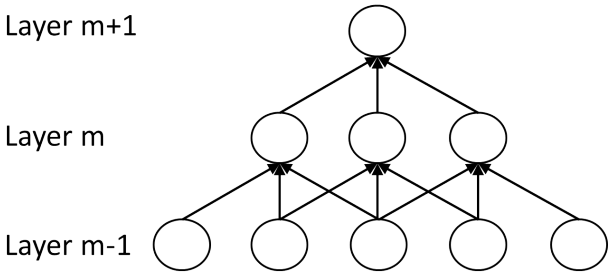
Figure 5.4: On the left side the gradient of a logistic activation function is illustrated and on the right side of the ReLU.

5.3 Convolutional neural networks (CNNs)

CNN were proposed by LeCun [87] and represent biologically-inspired variants of MLPs. This can be seen considering the visual cortex. It contains a complex arrangement of cells, where simple cells function as edge detectors, while complex cells provide invariance to the exact spatial position of the edges as well as their orientation. All cell types are characterized by *local receptive fields*, which act as local filters over their input domains and are tiled to cover the entire input space.

5.3.1 Architecture of CNNs

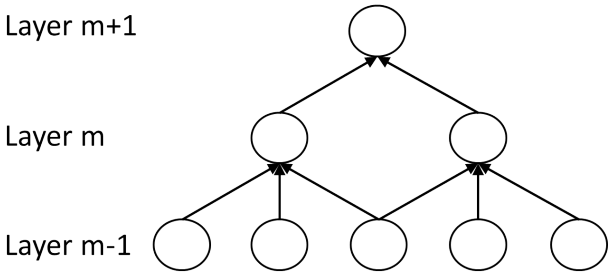
Fully connected networks suffer from the *curse of dimensionality*. With DNNs, the concept of localized receptive fields entails local connectivity in all hidden layers. The architecture thus ensures that the learnt filters produce the strongest response to a *spatially local* stimulus.



Local receptive fields also substantially reduce the number of connections to be learnt. However, stacking many such layers still would lead to global filters with many connections.

Weight sharing and translational invariance

Further reduction of the number of connections can be achieved via sharing corresponding weights.



Note that *weight sharing* resembles *convolution*:

- 1D-convolution

$$o[n] = f[n] \otimes g[n] = \sum_{u=-\infty}^{\infty} f[u]g[n-u] = \sum_{u=-\infty}^{\infty} f[n-u]g[u] \quad (5.1)$$

- 2D-convolution

$$o[m,n] = f[m,n] \otimes g[m,n] = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} f[u,v]g[m-u,n-v] \quad (5.2)$$

Striding each filter $\mathbf{h}_i$ across the entire input space allows for features to be detected regardless of their position in the input space. All features share the same set of parameters and combine into a feature map $\mathbf{H}^{(k)}$

$$h_{ij}^{(k)} = \tanh\left((\mathbf{W}^{(k)} \otimes \mathbf{x})_{ij} + b_k\right) \tag{5.3}$$

Training still employs modified backpropagation learning, whereby the gradient of a shared weight is simply the sum of the gradients of the parameters being shared. Weight sharing is also often accompanied with a max-pooling operation, which represents a nonlinear sub-sampling method. If max-pooling is employed, the outputs of the max-pooling layer are invariant to shifts in the inputs. Thus it provides additional robustness to spatial position and reduces the dimensionality of intermediate representations. Of course, the *max operation* can be replaced by an *averaging operation*.

Brightness invariance

CNNs commonly also contain *Local Contrast Normalization* (LCN) layers, where z-scoring is applied, i. e. in the LCN layer the mean is subtracted from the outputs of the max-pooling layer, and a division by the standard deviation of the input activities is performed. Such LCN layers provide, for example, brightness invariance, which is useful for image processing.

Modified BackProp

The standard error back-propagation algorithm needs to be modified to deal with deep architectures [87]: The forward pass performs an explicit computation with all weights. Thereby one has to remember which branch yields the max value. In the backward pass, all gradients $\partial J(\mathbf{W})/\partial w_m$ are computed for that branch only. When updating the weights, the average of all gradients from shared weights is used, for examples

$$\mathbf{w}_m - \eta \left(\frac{1}{M} \sum_{m'=1}^{M} \frac{\partial}{\partial \mathbf{w}_{m+m'}} J(\mathbf{W}) \right) \tag{5.4}$$

where $\mathbf{w}_m = \mathbf{w}_{m+m'}$ represent shared weights.

CNNs with multi-channel inputs

Images typically have multiple input channels, for example RGB channels. Thus we need to modify the filter to look at multiple channels. Weights are usually not shared across channels.

CNNs with multiple maps

Basic CNNs have only one filter per input position. This simple architecture can be extended to have many filters per spatial location. Each set of outputs, produced by a single filter, is called a feature map. To pass information forward, the output is treated as an image with multiple channels. Each output map represents an input channel. Thus a richer representation of the data can be achieved by each hidden layer having multiple feature maps. The weights $\mathbf{W}$ of a hidden layer can be represented in a 4D tensor containing elements for every combination of a destination feature map, a source feature map, a source vertical position, and a source horizontal position. The biases $\mathbf{b}$ can be represented as a vector containing one element for every destination feature map.

In the example given above, layer $m-1$ contains four feature maps and layer m contains two feature maps $\mathbf{H}^{(0)}, \mathbf{H}^{(1)}$. Pixel activities in $\mathbf{H}^{(0)}$ and $\mathbf{H}^{(1)}$ are computed from pixel activities of layer $(m-1)$ which fall within their 2×2 receptive field in the layer below. Thus the receptive field spans all four input feature maps. The weights $\mathbf{W}^{(0)}, \mathbf{W}^{(1)}$ of the feature maps $\mathbf{H}^{(0)}, \mathbf{H}^{(1)}$ form 3D weight tensors. Thereby, W_{ij}^{kl} denotes the weight connecting each pixel of the k-th feature map at layer m, with the pixel at coordinates (i, j) of the l-th feature map of layer $(m-1)$. Note that a randomly initialized filter acts very much like an edge detector. Also, weights are sampled randomly from a uniform distribution in the range $\{-\frac{1}{fan-in}, \frac{1}{fan-in}\}$.

5.4 U-Nets

CNNs are typically used for classification tasks, hence the above discussed dimension reduction steps do not matter. But image processing applications also need to preserve localization information. Also with biomedical image processing, a huge training data set is almost never available. One way around this bottleneck is to use a sliding window approach. However, the latter approach is very slow and highly redundant. Furthermore, localization accuracy has to be balanced against the use of context, and feature size is limited by window size. Especially if at the output spatial information is needed in form of a 2D array, like in a segmentation task, the dimension reduction steps need to be reverted by up-sampling. Ronneberger [129], to solve image segmentation problems, recently proposed a fully convolutional network called *U-Net* (see Fig. 5.5), which, at every up-sampling step, takes into account input from the corresponding down-sampling level to preserve information about spatial resolution at the various levels.

The fully CNN supplements a CNN by successive layers, where down-sampling operators (striding, pooling) are replaced by up-sampling operators. These sup-

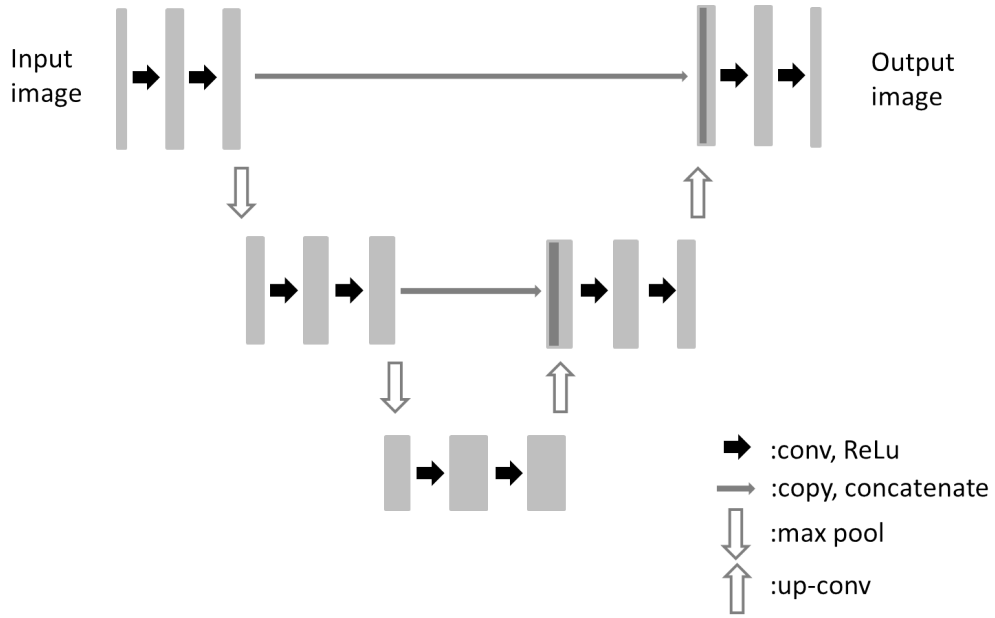


Figure 5.5: An illustration of a typical U-Net architecture.

plemental layers increase the resolution of the output. In addition, high resolution features from the contracting (down-sampling) paths are combined with upsampled output. A successive convolution layer can then assemble a more precise output. The up-sampling path contains a large number of feature channels. The latter allow the network to forward context information to higher resolution layers. Thus the *expansive* path and the *contractive* path run parallel to each other. This yields the *U-shaped* architecture.

CNNs commonly employ image patches rather than whole images. But a patch-based approach limits context information and feature size. The receptive fields of these networks are bounded by the patch size. U-Net CNNs allow to use entire images of arbitrary sizes, instead. This becomes important, if large structures need to be recognized like in a segmentation task or the prediction task studied in this thesis.

As no classification is intended, the U-Net doesn't have any fully connected layers. The output map only contains pixels whose full context is available in the input layer. Hence arbitrarily large images can be analyzed by an overlap-tile strategy. This tiling strategy prevents the resolution to be limited by the GPU memory rather than the patch size. With biomedical images usually only few traing data is available. Hence excessive data augmentation by elastic transformations is necessary. Such elastic transformations can be realistically simulated with biomedical

images. The U-Net simply learns invariance to such transformations.

5.4.1 Network Architecture

Following, a typical U-net architecture will be exemplified and its functioning described. Later this generative architecture will be taylored to the specific needs of the tasks to be solved. The network architecture used by Ronnenberg et al. [129] has a contracting (down-sampling) and an expanding (up-sampling) branch. It was designed for segmentation tasks. Other tasks may need different architectures as will be discussed later on.

- The *contracting path* of a typical U-Net looks like a CNN with
  - repeated unpadded or padded convolutions
  - a leaky ReLU
  - a max-pooling operation with stride 2, i. e. a kernel size 2×2, for down-sampling
  - at each max-pooling layer, the input image size is divided by the size of the max-pooling kernel size
  - at each down-sampling step, the number of feature maps is doubled
- The *expanding path* consists of
  - an up-sampling of the feature maps, implemented with a convolution kernel
  - kernel weights that are learned during training
  - an up-convolution that halves the number of feature maps
  - a concatenation with the related (cropped in case of no padding) feature map from the contractive path
    - this incorporates information available at the down-sampling stage into the up-sampling operations
    - hence fine-details captured in the descending part is used at the ascending part
  - convolutions followed by a leaky ReLU
  - The cropping accounts for the loss of border pixels in every unpadded convolution. In case of padding, cropping is not needed.
- The final convolution layer maps each feature vector to the desired number of tissue classes to be segmented.

During *up-sampling*,

- the output of any previous step is convolved
  - by initially up-scaling the image by a factor of two using nearest neighbor interpolation
  - then convolving with a convolution kernel size 2×2 and
  - followed by a leaky ReLU layer
- Then concatenate with the corresponding down-sampling layer.
- Finally, two convolution layers, each followed by a leaky ReLU layer, are applied to this concatenated image.

After up-sampling, a final convolution layer and following sigmoidal unit produce probability outputs for each tissue class. Convolutions can be padded or not. A padded convolution allows the entire image to be processed at once.

5.4.2 Training

Following, network training will be discussed from a general perspective assuming a biomedical image segmentation task, for example kidney segmentation in an X-ray CT image slice. Actually performed training procedures will be discussed later when it comes to discuss specific applications. For training U-nets, input images and related output maps are commonly used. Learning the kernel weights is performed with SGD learning. The SGD can be implemented with Caffe, Keras and Tensorflow or PyTorch. In this study, Tensorflow and Keras have been employed. Also to preserve input image size, zero-padding has been applied. Finally, large input tiles, or alternatively more convolutional layers, should be favoured in general over a large mini-batch size. A high momentum, furthermore, will assure that a large number of previously seen training samples determines the update in every subsequent optimization step.

The loss function Q is evaluated by first applying a *softmax* transformation pixelwise to the final feature map yielding

$$p_k(\mathbf{x}) = \frac{\exp(a_k(\mathbf{x}))}{\sum_{k'=1} \exp(a_{k'}(\mathbf{x}))} \tag{5.5}$$

where

- $a_k(\mathbf{x})$ denotes the activation in feature channel k at the pixel position $\mathbf{x} \in \Omega \subset \mathbb{Z}$
- K denotes the number of tissue classes

- $p_k(\mathbf{x})$ is the approximated maximum-function
  - Hence $p_k(\mathbf{x}) \approx 1$ for the k that has the maximum activation $a_k(\mathbf{x})$
  - And $p_k(\mathbf{x}) \approx 0$ for all other k

The softmax operation is commonly combined with a *cross-entropy loss function* $Q(p|w)$. The latter penalizes at each spatial position $\mathbf{x} \in \Omega$ the deviation of the class probability $p_{l(\mathbf{x})}(\mathbf{x})$ from 1 using

$$\begin{aligned} Q(p|w) &= \sum_{\mathbf{x}\in\Omega} w(\mathbf{x}) \log(p_{l(\mathbf{x})}(\mathbf{x})) = KL(p|w) + H(w) \qquad (5.6) \\ KL(p|w) &= \sum_{\mathbf{x}\in\Omega} w(\mathbf{x}) \log\left(\frac{p_{l(\mathbf{x})}(\mathbf{x})}{w(\mathbf{x})}\right) \\ H(w) &= \sum_{\mathbf{x}\in\Omega} w(\mathbf{x}) \log(w(\mathbf{x})) \end{aligned}$$

Here $KL(p|w)$ denotes the Kullback-Leibler divergence (*Kullback-Leibler divergence* (KL)) and H_w the information entropy of the distribution $w(\mathbf{x})$. Furthermore, $l : \Omega \to \{1, \ldots, K\}$ is the true label of each pixel, while $w : \Omega \to \mathbb{R}$ is a weight map, which can be considered the true distribution $w(\mathbf{x}) \equiv p(\mathbf{x})$. The latter needs to be pre-computed for each ground truth segmentation. For example, Ronneberger et al. [129] considered cell segmentation, where these weights compensated the different frequency of pixels from a certain tissue class. For cell segmentation, the latter class encompassed pixels in the small borders between close cells. The separation border was computed using morphological operations and the weight map was computed as

$$w(\mathbf{x}) = w_c(\mathbf{x}) + w_0 \cdot \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right)$$

Here $w_c : \Omega \to \mathbb{R}$ denoted the weight map to balance the class frequencies, $d_1 : \Omega \to \mathbb{R}$ denoted the distance to the border of the nearest cell, $d_2 : \Omega \to \mathbb{R}$ denoted the distance to the border of the second nearest cell and w_0, σ had to be set by the user.

Good initialization is extremely important. Weights should be initialized such that each feature map in the network has approximately unit variance. In a deep network with alternating convolution and ReLU layers, weights should be drawn from a Gaussian distribution with variance $\sigma^2 = 2/N$. Here N denotes the fan-in of a node.

Data augmentation Data augmentation is necessary if only few samples are available. It helps to teach the network desired invariances and robustness properties. With biomedical images, mostly shift and rotation invariance is needed, but also robustness to deformations and robustness to gray value variations need to be achieved. Random elastic deformations seem best for data augmentation with only few annotated samples. Drop-out layers at the end of the contracting path implicitly augment data also. In this study data augmentation is not used, however.

Chapter 6

Dataset

6.1 Patient collective

Table 6.1: Patient data concerning the DOTATOC radiotherapy. Here m denotes the weight of each person, A_0 the initially injected radioactivity, g the grading of the NET, CgA_1 denotes the CgA before therapy and CgA_2 after therapy, Δt the time betweent the CgA values, n the number of therapy cycles and the sex is coded in 1 for female and 0 for male.

Age $[y]$	sex	g	m (kg)	A_0 (MBq)	CgA_1 $(\mu g/l)$	CgA_2 $(\mu g/l)$	Δt (m)	n
45	1	3	64	6077	494	1026	5	4
66	0	3	79	6375	1121	706	4	4
54	0	1	77	7054	661	1956	4	2
67	1	2	70	6821	1352	1473	6	3
78	0	2	88	5773	34	31	4	3
54	0	2	56	7023	54	77	4	3
67	0	2	75	5809	597	1748	2	2
54	0	2	71	6644	1319	682	4	4
77	1	2	49	6378	224	246	1	1
52	0	3	78	6233	46	83	5	2
52	0	3	103	7265	76	47	4	3
71	0	2	118	6757	51	51	3	2
49	0	3	80	6328	139	214	2	2
66.5		2.3	77.5	6503	589.8	768.1	3.9	2.8

This study comprises 26 patients suffering from NETs (13) or PC (13) and

Table 6.2: Patient data concerning the PSMA radiotherapy. Here m denotes the weight of each person, A_0 the initially injected radioactivity, $Gleason$ the Gleason Score of the PC, PSA_1 denotes the PSA level before therapy and PSA_2 after therapy, Δt the time betweent the PSA values and n the number of therapy cycles.

Age $[y]$	Gleason	m (kg)	A_0 (MBq)	PSA_1 (ng/ml)	PSA_2 (ng/ml)	Δt (m)	n
80	9	88	6699	1277	594	2	1
67	8	104	6997	96	120	6	2
75	9	74	6314	22	5	2	1
55	6	113	5931	122	523	2	1
61	9	85	6899	32	74	5	2
66	9	97	7139	92	199	2	1
63	9	94	7103	484	289	2	1
65	9	80	7030	58	0.1	3	2
66	9	86	6545	469	445	1	1
78	9	106	6768	93	86	3	2
79	9	75	6915	138	25	2	1
62	8	68	6824	2251	1946	2	1
55	9	94	7006	148	53	2	2
65.7	8.4	90	6782	406.3	336.0	2.6	1.4

which underwent a ^{177}Lu-DOTATOC or ^{177}Lu-PSMA therapy. The patient cohort consists of 23 male and 3 female patients with an average age of 63.8 ± 10 years at time of therapy. The on average injected activity was $A_0 = 6643 \pm 421$ MBq.

The information about the patients suffering on NET is collected in table 6.1 and for the patients with prostate cancer in 6.2.

6.2 Image Acquisition

SPECT: Data was acquired on a SPECT/CT system (Siemens Symbia T2) at time points 4h, 24h, 48h, and 72h after administering the radiopharmaceutical. The acquisition of the 24 h p.i. image was carried out as part of a hybrid SPECT/CT acquisition, After manual co-registration, the CT part was used for attenuation correction of the subsequent SPECT. Acquisition was done based on an in-house standard quantitative ^{177}Lu - protocol, which is described in detail in [132]. Therefore, it is herein only briefly outlined:

- SPECT using medium energy collimators, 3° angular sampling, 15 min total dwell time

- Iterative ordered-subset conjugate-gradient reconstruction of the 208 keV photopeak data with 24 iterations, 1 subset, matrix 256x256
- Point-spread-function modelling in reconstruction
- Triple energy window based scatter correction
- CT-based attenuation correction
- No post-reconstruction smoothing

The reconstruction, using the ordered-subset conjugate-gradient algorithm, was carried out on a Siemens research workstation. The algorithm outputs fully quantitative reconstructed SPECT images providing activities per voxel measured in $[Bq/ml]$. Furthermore the SPECT data is alined with the corresponding CT data.

CT: Additionally, a low-dose CT was carried out as part of the multi-modal SPECT/CT acquisition to enable necessary attenuation corrections of the SPECT images. The CT covered the same field-of-view as the SPECT and was acquired and reconstructed using the following parameters:

- Slice collimation of 2 x 5 mm, pitch of 1.8, time per rotation of 0.8 s, tube voltage of 130 kVp, tube current of 30 mAs effective
- Filtered-Back-Projection reconstruction with B08s and B41s Kernels, 512x512 matrix, 2.5 mm slice thickness
- B08s image was used for attenuation correction of the SPECT data
- B41s image was used for defining organ and tumor volumes-of-interest

Reconstructed SPECT images were partitioned into voxels of size $V_{vox} = 4.79\ mm^3$ and were cropped to 82^3 voxels. Thereby large portions of the reconstructed volume were removed which predominantly contained air.

6.3 Common Preprocessing

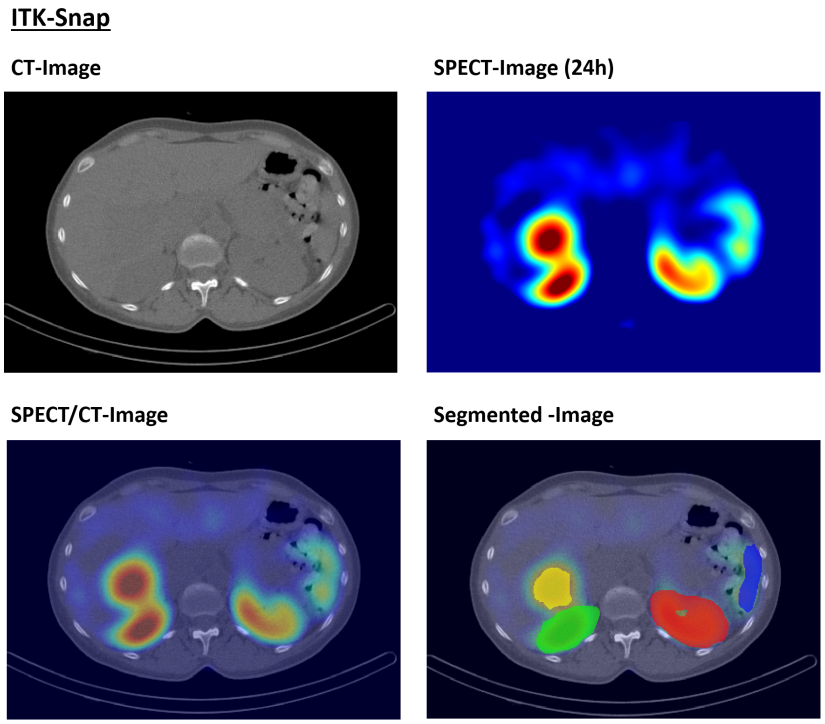


Figure 6.1: CT-image, SPECT-image and hyprid SPECT/CT-image as shown in ITK-snap. In the right bottom corner the segmentation of the right and left kidney, the spleen and a liver tumor can be seen.

In figure 6.1 a CT-image, a SPECT-image and a SPECT/CT-image are presented. For all patients, the kidneys, spleen and tumors were segmented by a medical expert. The tumors were sorted to bone lesion, lymph nodes, visceral and primary tumor. The segmentation was done using the software ITK-snap [173], which was carried out on the original CT data sets. The resulting segmentation was subsequently down-sampled to match the voxel-size of $(4.7mm)^3$ of the SPECT image. For each patient, the following information is available:

- SPECT-Image acquired 4h after injection
- SPECT-Image acquired 24h after injection
- SPECT-Image acquired 48h after injection
- SPECT-Image acquired 72h after injection

- CT-Image acquired 24h after injection
- Organ and tumor segmentation

Exemplarily, for one patient and one slice the eight images are illustrated in figure 6.2.

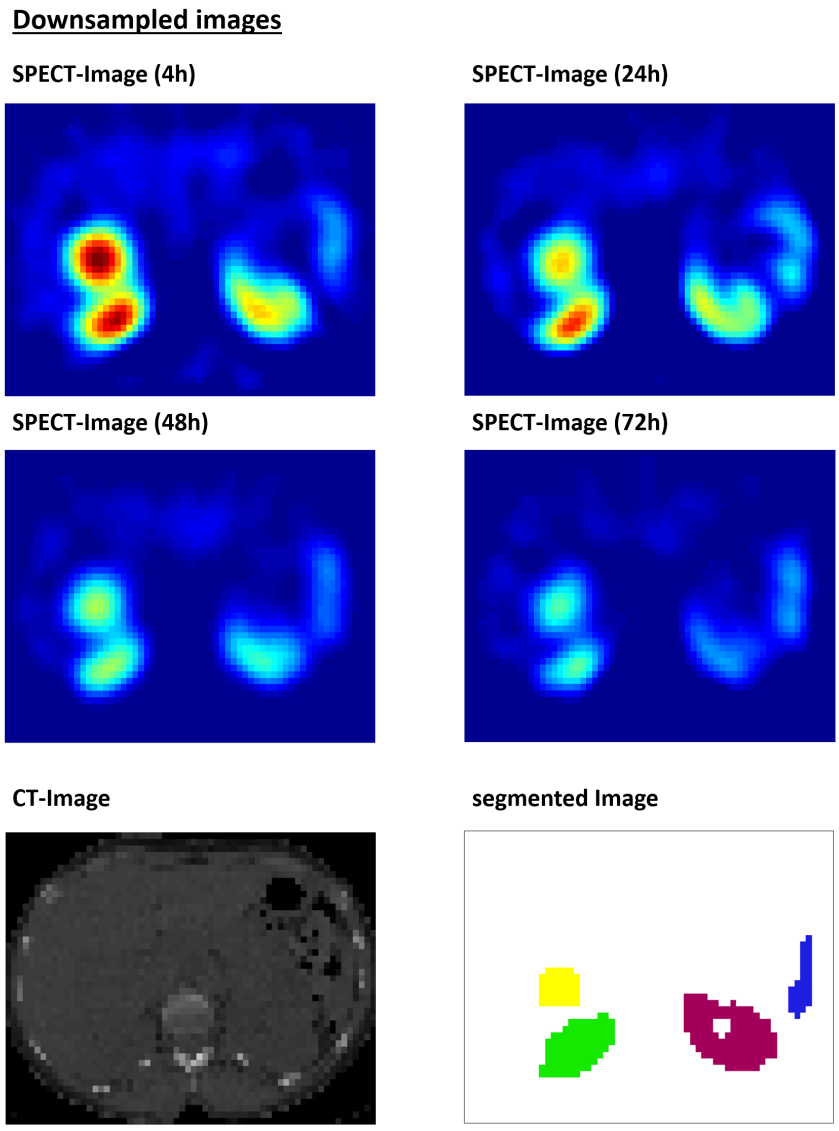


Figure 6.2: The four SPECT-images, the CT-image and the corresponding segmentation map illustrated for one slice of one patient.

Chapter 7

Error of the standard MIRD method

This chapter is abstracted from [TIG-1]. In this chapter, the influences on the dose value calculated by the standard MIRD method as described in chapter 1.2.4 were evaluated. In figure 1.1 the basic steps of this method are illustrated. The following steps were examined:

- Segmentation of an organ on CT,
- Registration of 4 SPECT-images on CT,
- Amount of activity within organ,
- Integration of time-activity-curve,
- Scale with standard phantom and
- Multiplication with S-Value.

7.1 Segmentation of an organ on CT

The kidneys of 26 patients were segmented by two medical physicists. The kidneys deliberately were not segmented very precisely, because in clinical routine this approach is very time consuming. Furthermore, the renal pelvis was included by some patients and also kidney cysts.

7.2 Registration of 4 SPECT-images on CT

The SPECT images from different time points were registered by a medical physicist to the CT acquired roughly $24\ h$ after injection of the radiopharmaceutical.

Afterwards the organ-specific activity is calculated which varies between patients. Two medical physicists registered the 4 images of 26 patients to the corresponding CT and calculated the time activity curve where each time point is weighted equally. The activity contour remained unchanged.

7.3 Amount of activity within organ

Normally the activity contour is defined by a medical physicist. The activity contained in the kidney is spread because of the partial volume effect. Therefore the activity contour is defined with some margin around the kidney. Kidneys with tumors and joint organs are excluded. Consequently, the actual contour depends on the defining medical physicist. The influence of this "human factor" onto the accuracy will be evaluated in this section. Two medical physicists defined proper ROIs in equally registered SPECT/CT images. Afterwards, a simple exponential time activity curve with equally weighted activities was approximated and integrated.

7.4 Integration of time-activity-curve (TAC)

Usually a SPECT image is acquired at four time points ($4\ h$, $24\ h$, $48\ h$, $72\ h$). Two patients were excluded in this section because less than four images were acquired. In most cases, the measured activity in the kidneys of the second time point is higher than for the first one. A reason for this apparent anomaly could be a delayed onset of physiological processes e.g. delayed tracer excretion from the kidneys. If so, the later time points seem more important for the determination of the time activity curve than the activity at the first time point [85]. One possibility to take this into account would be a weighted fit, as it is practiced in this study. For 25 patients, the time-activity curve of each patient was fit to a mono-exponential function. The time integrated function yields the *Area Under the Curve* (AUC). The weights were chosen as shown in table 7.1.

Another possibility to manipulate the fit is to add a time point at $600\ h$ with an activity value of zero as proposed in [69]. Strictly speaking, this contradicts the exponential character of the decay curve, but in practice the effect is to force the fit curve to zero.

A mean AUC per patient has been calculated from the measured AUCs of all patients as well as a related mean relative standard deviation. Furthermore, the mean relative deviation and the maximal and the minimal deviations were also determined per patient. Finally, the mean over all patients of these quantities were calculated.

Table 7.1: Weights used for fitting a time activity curve

time point	4h	24h	48h	72h
weights x	1	1	1	1
weights a	1	10	10	10
weights b	1	100	100	100
weights c	1	1	10	10
weights d	1	1	100	100
weights e	1	1	1	10
weights f	1	1	1	100

For four patients of the data set, an additional time point at $144\ h$ was acquired. If this point is taken into account, the same parameters of best fit are obtained as for the inclusion of an additional point at $t = 600\ h$, as their weights are assumed to be equal.

7.5 Scale standard phantom and multiply with S-Value

To calculate, from the number of decays per organ, the corresponding absorbed dose, the toolbox OLINDA uses S-values calculated by MCs on a standard phantom. But for the MC, a homogeneous activity distribution is assumed. To get closer to the real patient, the S-value was scaled with the inverse organ mass m_T. To evaluate the influence of the assumption of a standard phantom, MCs of the whole patient with an inhomogeneous but known activity distribution were performed. Additionally, the number of nuclear disintegrations per organ from the isotope distribution was taken and multiplied with the S-value and the inverse organ mass.

7.6 Results

Dependence of estimated accumulated activity on various influencing factors and related error statistics are collected in the following table.

The components of the total error of the dosimetry chain are listed in table 7.3.

Table 7.2: Dependence of estimated accumulated activity on various influencing factors and related error statistics

Calculation of $\tilde{A}(\mathbf{r}_S)$	$\langle\tilde{A}(\mathbf{r}_S)\rangle$	Error (%)
SPECT/CT registration	$3.28\ h$	$3.4 \pm 4.4\ (0.02 - 13.9)$
Activity contours	$3.33\ h$	$6.5 \pm 7.5\ (0.3 - 28.6)$
Time activity curve	$2.91\ h$	$7.2 \pm 4.8\ (1.6 - 22.9)$

Table 7.3: Total error estimate

Dosimetry step	Error (%)
Resulting AUC error	17.1 ± 16.7
Dose kernel vs. Monte Carlo	16.7 ± 12.8
Kidney segmentation	5.4 ± 2.9
Activimeter accuracy	5.0
Total Error	25.0

7.7 Discussion

The absorbed dose for an organ of interest and its related uncertainty can be calculated by the following expressions:

$$
\begin{aligned}
D(\tilde{A}, A_0, m_{Pat}) &= \tilde{A}(\mathbf{r}_T) \cdot \frac{S_{phant} \cdot m_T^{Phant}}{m_T^{Pat}} \cdot A_0(\mathbf{r}_T, 0) \qquad (7.1)\\
dD(\tilde{A}, A_0, m_{Pat}) &= \sqrt{\left(\frac{\partial D}{\partial \tilde{A}} d\tilde{A}\right)^2 + \left(\frac{\partial D}{\partial m_{Pat}} dm_{Pat}\right)^2 + \left(\frac{\partial D}{\partial A_0} dA_0\right)^2}\\
\frac{dD(\tilde{A}, A_0, m_{Pat})}{D(\tilde{A}, A_0, m_{Pat})} &= \sqrt{\left(\frac{d\tilde{A}}{\tilde{A}}\right)^2 + \left(\frac{dA_0}{A_0}\right)^2 + \left(-\frac{dm_{Pat}}{m_{Pat}}\right)^2}
\end{aligned}
$$

where D is the dose in Gy, $\tilde{A}$ the area under the TAC, S_{phant} the S-value for a standard phantom, m_{phant} the targeted organ mass of the phantom, m_{pat} the corresponding targeted organ mass of the patient and A_0 the applicated activity. The total dose error is estimated, assuming Gaussian error statistics, from the known uncertainties of the independent variables.

All these quantities entering an estimate of the absorbed dose have been evaluated previously and are reported variously in literature. The residence time is influenced by the registration, activity contour and the weights of the fit function for the time-activity curve. The mean deviation caused by human registration is

$3.4 \pm 4.4\%$. In [133] different activity contours in both planar and 3-dimensional data arrays were evaluated and differences were asserted. In the current study, the relative deviation between the AUC, calculated independently by two physicists, was on average $6.5\% \pm 7.5\%$. If the contour is drawn by a human, any tumor located close to the kidneys can be excluded deliberately.

The weights for the fit function also influence the AUC for dose calculation. In [45] different fit functions were chosen for planar images. In the current study, however, the weights were chosen differently, whereby always the later images became stronger weights. On average across the whole patient cohort, a mean relative deviation of $7.2\% \pm 4.8\%$ between fit functions with different weights was achieved. The deviation strongly depended on the activity values which have been calculated by summing up all activity values within the activity contour. To extract a deviation value for the residence time, one can add the individual errors as $(3.4 \pm 4.4\%) + (6.5\% \pm 7.5\%) + (7.2\% \pm 4.8\%) = 17.1\% \pm 16.7\%$. The range of errors which can happen is between 0.4% and $33,8\%$, which strongly depends on the patient anatomy, position of the tumor and the registration of different SPECT images.

After the residence time is estimated, the value is multiplied by the organ specific S-value. The S-value has been scaled by the organ mass to improve the results. A mean deviation, averaged over all patients, of $16.7\% \pm 12.8\%$ results between a dose value from a full Monte-Carlo simulation and one estimated by the scaled S-values. The values are close to those calculated by [80], where a mean deviation of 26% in 9 patients was obtained. For the evaluation, the individual total organ activity as well as the mass of the organ were known. In general, the mass of the segmented kidney has to be estimated from a CT image. For this reason the interobserver variability of the kidney segmentation was evaluated. An averaged precision of the kidney volume of $5.4\% \pm 2.9\%$ between two observer was achieved. It is close to the result from [9] where a deviation of 11.2% between two observers was reported.

The product of scaled S-value and AUC is multiplied with the applicated activity. This value can be measured by an activimeter with a accuracy of 5.0%.

All in all a relative error of $\sqrt{(17.1\%)^2 + (16.7\%)^2 + (5.4\%)^2 + (5.0\%)^2} = 25.0\%$ has to be expected during the dosimetry within a ^{177}Lu radiotherapy. It is important to take into account that some of the relative errors which can happen cancel each other. A good way to avoid errors is a standardized method where all patients are handled according to a prescribed protocol.

-

Chapter 8

Time-integrated activity map (TIA)

This chapter is abstracted from [TIG-2]. In this chapter the four SPECT-images as shown in figure 6.2 will be integrated voxel-wise to obtain the number of decays per voxel. The activity values will be denoised by the particle filter. After the denoising, a mono-exponential function is fit to the data and integrated over time.

8.1 Model equations

For dosimetry it is more important to have a patient specific model than an on-line tracking method. Therefore all SPECT images were used to estimate a state evolution model. First of all, VOIs were segmented in a CT image by a medical expert. To this contour a safety margin of 2 voxels was added to account for partial volume effects as well as deviations from the registration or inter-organ movements. By summing up the activity values of all voxels enclosed by the contour, an *organ-specific total activity* was obtained for each time point. An additional time point was added at $t_k = 600\ h$, at which the specific activity was set to $0\ Bq/ml$ deliberately. The time-dependent activity (see figure 8.2) was expressed by a parametrized analytic function which was fit to the observed time-dependent activities according to

$$\ln\left(\frac{a(t)}{a(0)}\right) = -\left(\frac{ln(2)}{T_{1/2}} \cdot t\right) = -\frac{t}{\tau_{eff}} \tag{8.1}$$

Note that the effective lifetime

$$\tau_{eff} = \left(\frac{1}{\tau_{phys}} + \frac{1}{\tau_{biol}}\right)^{-1} = \frac{T_{1/2}}{\ln 2} \tag{8.2}$$

is composed of the physical lifetime of the radionuclide and its biological lifetime due to the metabolic turnover of the radio-pharmaceutical. While the biological half-life of the isotope should be identical across an organ, the amplitude

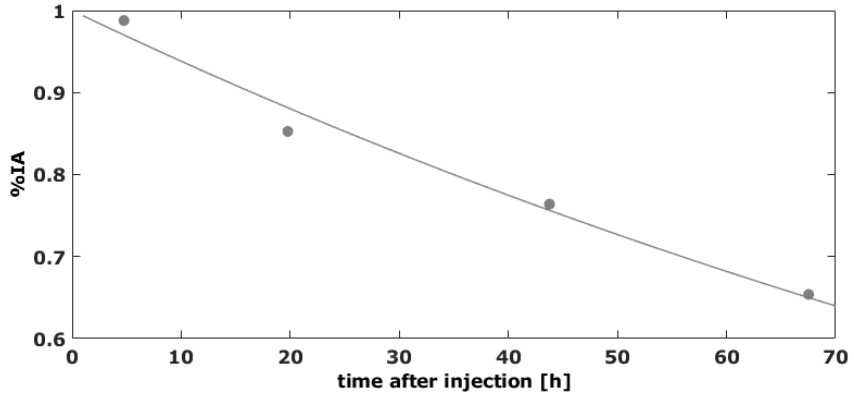


Figure 8.1: Time activity curve for one kidney: $f(0) = 1\%IA, T_{1/2} = 77.9h$.

$a(0)$ of the TAC is different for every voxel. Note that $a(0)$ is different from the initially injected activity A_0 and generally unkown.

State evolution model (SEM): For each voxel of the kidney an own *State Evolution Model* (SEM) is computed. At the beginning, the four activity values of the same voxel plus an additional value of $a(t_k) = 0Bq/ml$ at $t_k = 600\ h$ are approximated by a mono-exponential decay. Hereby, the half-life is assumed to be equal to the previously computed half-life of the whole kidney. From this fit, the amplitude $\langle a(0) \rangle$ of the state evolution model can be obtained. Moreover, the inevitable noise on the data is modeled as Gaussian white noise with an amplitude corresponding to $0.1\langle a(0) \rangle$.

Observation model (OM): The *Observation Model* (OM) $\mathbf{h}$ represents the distribution of acquisition times of the SPECT image and is modeled by a Normal distribution around the estimated particle.

Similarity measure The aim of the voxel-wise integration of the activity is to obtain the number of nuclear disintegrations N_n^{Lu} for each voxel $n \in S$ according to:

$$N_n^{Lu} = \int_0^{t_K} a_n(t')dt' \approx \sum_{k=1}^{K} a_n(t_k) \cdot \Delta t \tag{8.3}$$

where $K = 5$ in this study. The activity distribution is assumed to be identical for each voxel, hence it can be estimated for each organ at every time point. This fact can be explored to find a good parameter configuration. As similarity measure the

KL was used. An averaged histogram of the $k = 4$ time points was calculated, for which the histograms were centered. A histogram of the voxel activities $h(\langle a_n \rangle)$ was calculated, where $\langle a \rangle_n = (1/K) \sum_k a_n(t_k)$ denotes the activity of voxel n averaged over the four time points. The similarity between different histograms has been computed employing the KL.

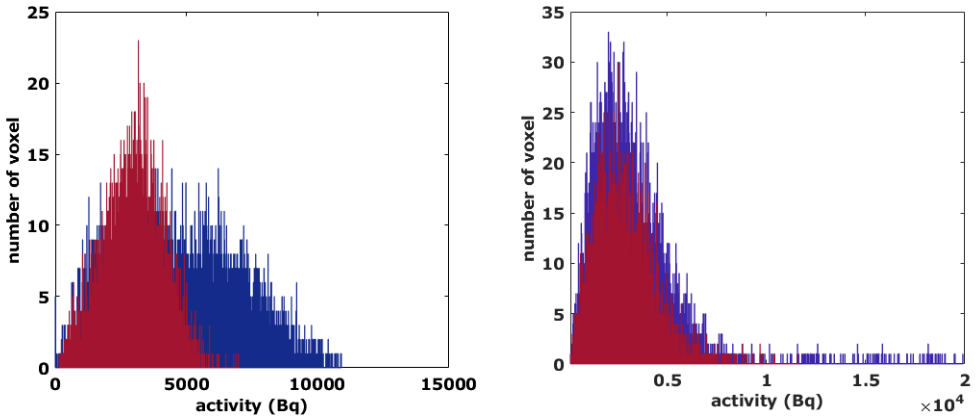


Figure 8.2: Left side two histograms with a Kullback-Leibler divergence of 0.67 and on the right side with 0.12.

8.2 Other Methods

Instead of fitting the activity values as smoothed by the particle filter ('particle fit'), one can use another approach to calculate a voxel-wise decay map. One can fit a mono-exponential function to the four activity values per voxel. In the following this method is called 'simple fit'.

For each method the TIA in a specific source region S was calculated by integrating, for every voxel, the fit function over time and sum over all voxels.

$$TIA_S = A_0 \sum_{n \in S} \int_0^\infty \cdot e^{-\frac{t}{\tau_{eff,n}}} dt$$

where S is the source region, n indicates one voxel, A_0 denotes the injected activity and $\tau_{eff,n}$ the half-life, which can differ between the voxels for both, the 'simple fit' and 'particle fit' method. Note that the approximation $a_n(t_0) \approx A_0 \forall\ n \in S$ accounts for the fact that the complex metabolic processes which distribute the radioactive pharmaceutical in the body is generally unknown. Finally, also the distribution of half-lives across all voxels in the VOIs was calculated.

8.3 Results

8.3.1 Parameter evaluation

Finding a good parameter configuration, which yields similar histograms of the number of decays within an organ, and which also yields a good match to the measured activity distribution of all activity values within a VOI, the standard deviations for the state evolution model and the observation model were considered adjustable parameters. The configurations were computed for all 26 patients, and for each configuration, the histogram of the number of decays per voxel was determined and compared with the averaged activity histogram. The KL was computed for all configurations. The σ_{obs} was varied between $0.35 \leq \sigma_{obs} \leq 0.99$. The considered standard deviations are rather large because the estimated precision of the activity measurement through SPECT imaging is in this range according to [128].

In the following table, the KLs, averaged over all patients, are collected:

Table 8.1: Kullback - Leibler divergences, averaged over the entire patient cohort

model$\diagdown$ observation	0.35	0.5	0.75	0.99
0.01	0.12	0.17	0.17	0.17
0.10	0.16	0.17	0.17	0.17
0.25	**0.16**	0.16	0.16	0.16
0.35	0.58	0.54	0.48	0.46

The values in each row are very similar to each other, indicating that the KL is rather robust against uncertainties in the observation model. The smallest value for the KL is marked in blue.

Table 8.2: Relative deviations of the sum of voxel-wise integrated activities from the corresponding whole organ integrated activity in percent

model $\diagdown$ observation	0.35	0.5	0.75	0.99
0.01	-34.2%	-32.6%	-32.6%	$\mathbf{-32.6}\%$
0.10	-35.8%	-33.7%	-33.4%	-33.4%
0.25	-44.4%	-40.3%	-39.0%	-38.5%
0.35	-64.4%	-59.3%	-56.3%	-55.9%

As a second decision criterion for an optimal parameter configuration, the relative deviation of the total number of decays estimated either by integrating the

time-dependent activity at every voxel and then summing over all voxels within the VOI or by integrating the whole organ (here the kidney) activity over time was calculated.

In the following table the resulting relative deviations are collected. The deviations for the total number of decays are large because of a high statistical uncertainty in the voxel activity values. To correct for these large deviations, the voxelwise integrated activity map can be normalized with the total number of decays per organ. For the following results, we choose the parameters as $\sigma_{observation} = 0.35$ and $\sigma_{model} = 0.10$.

To obtain stable results for the single activity values, the number of particles for each time step was varied. In figure 8.3 the determined difference between the total number of decays per organ for different numbers of particles and the total number of decays for 2000 particles is illustrated. With only 150 particles, the difference to the reference value almost disappears.

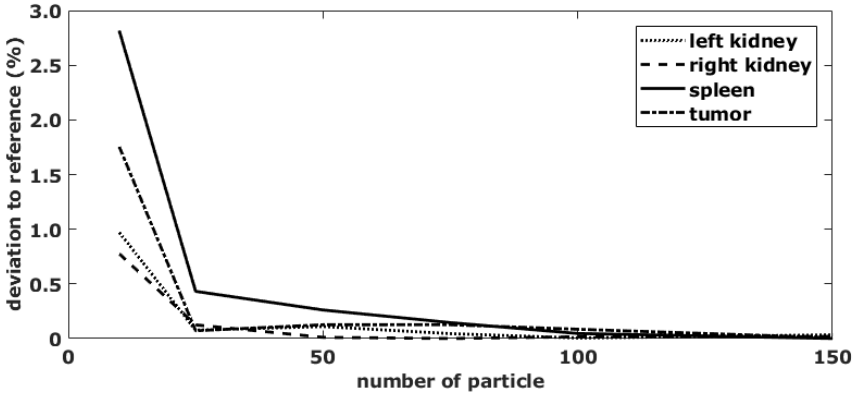


Figure 8.3: Variation of the number of particles.

Moreover, while applying the particle filter, different resampling schemes were evaluated, but there was no difference in the total number of decays per organ. Therefore we decided to perform no resampling. The possibility to loose too much diversity is very low with only four time steps.

8.3.2 Voxelwise TAC

For each voxel, a state evolution model was defined through four measured time points. According to the state evolution model and the observation model, particles were generated and an estimation of the true activity value was obtained through its likelihood. An analytic function was then adapted to either the four estimates of the true activity values or the observed activity values. In figure 8.4, the

actual measurements and the estimates of the state evolution model as represented by properly adapted analytic functions, whereby the adaptation to the observed values is drawn in black and the adaptation to the estimates of the true activity values is colored in gray. Through particle filtering, the outlier at the second time point could be eliminated very well.

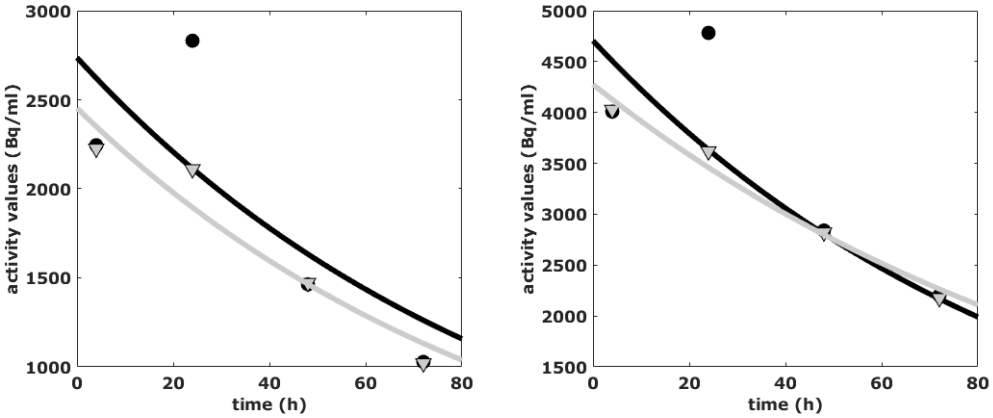


Figure 8.4: Time activity curve vor two different voxels at four time steps.

8.3.3 Whole patient

In figure 6.2, one slice of the SPECT images at four time points for one patient have been illustrated. Now the voxel-wise results for the integration of the TACs are shown in figure 8.5. The red voxels in the right image, showing the results from the simple fit method, are those with a diverging integral, the calculated number of decays is equal to infinity.

The voxel-wise integration of the particle filtered activity values was performed for the whole patient. For the segmented organs (kidney, spleen and tumor), a corresponding half-life was calculated and used for the state evolution model. For the voxels outside the VOIs, an over all voxels from the same tissue class averaged half-life was calculated and used for the model. For each VOI, a histogram of the number of decays per voxel was obtained and is shown in figure 8.6. A second histogram is also shown, were the number of decays per voxel is presented, which is calculated by a simple voxel-by-voxel integration. Both histograms are almost indistinguishable.

Considering the histograms of all voxel-specific half-lives, which were used in the analytic function for the determination of the TIA, the particle filter method yielded histograms centered at the half-life from the state evolution model, while

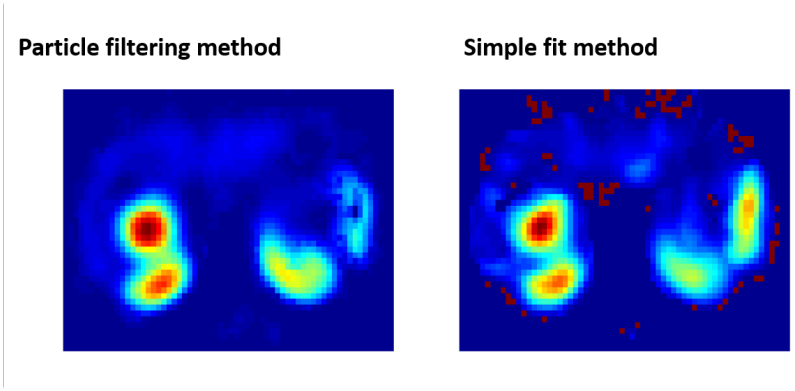


Figure 8.5: One slice of the results for the particle filtering method (left) and the simple fit method (right).

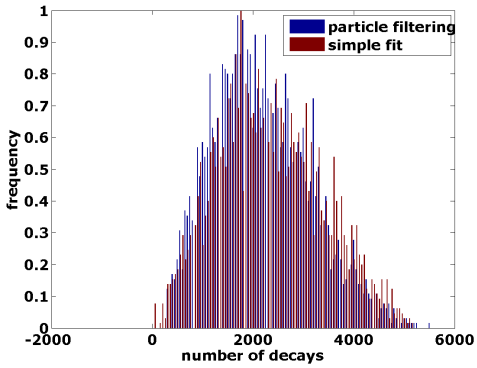


Figure 8.6: Histograms of number of decays per voxel for the particle filter method (blue) and the voxel-wise simple fit of the four activity values (red).

the simple fit method very often yielded half-lives larger than the physical half-life of Lutetium (see figure 8.7). Also the widths of the two distributions differ considerably. The much narrower width of the distribution for the particle filter method depends on the σ parameters of the SEM and the OM.

8.4 Discussion

In this work, we proposed a method for processing voxel-wise TACs prior to fitting them to mono-exponentials and subsequently integrating them temporally. The TACs were obtained by sequential, quantitative SPECT/CT imaging for dosimetry of 26 patients under Lu-177 targeted radiotherapy. The proposed pre-processing method is a *Particle Filter*, which belongs to the class of sequential Monte-Carlo

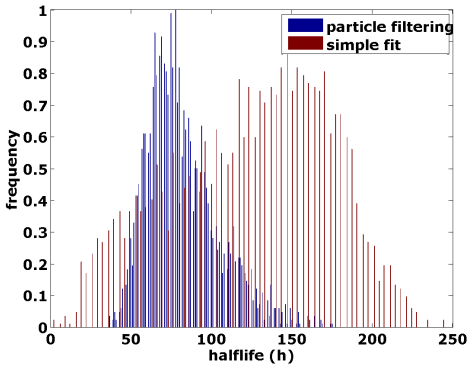


Figure 8.7: Histograms of halflives per voxel for the particle filter method (blue) and the voxel-wise simple fit of the four activity values (red). The halflife for the whole organ is $81.6h$.

methods. The aim of the method is a de-noising of the observed TACs. The goals were the following:

- The whole-organ TIA map, as obtained on basis of the filtered data, should be as similar as possible to the conventional evaluation. The latter obtains TIA maps by fitting and integrating the organ-averaged time-activity curve.
- The histograms of the TIA maps, resulting from the PF, should resemble the TIA-histograms, as initially estimated from SPECT images.

We have shown that *Particle Filters* are suitable for a reduction in the variance of the effective half-lives of voxels in kidneys, spleen, and tumors. Most notably, the number of voxels which exhibit half-lives longer than the physical half-life of the radioisotope is significantly reduced.

For two voxel-specific methods to estimate whole organ TIAs, namely *particle fit* (PFF) and *simple fit* (SF), we quantified deviations to whole-organ TIAs estimated by time-integration of the whole organ activity.

When comparing the resulting half-lives, we found that application of particle filtering results in a strong reduction of variance, compared to SF. As intended, the voxel-specific half-lives for PFF were distributed around the half-life used in the state evolution model. Importantly, the number of voxels which exhibit a longer-than-physical half-life was greatly reduced with PFF, compared to SF. In summary, the PFF method does not fix the half-life to a single value, rather it exhibits a narrow distribution and reduces the number of implausible half-lives to almost zero in our study.

Unfortunately, literature for comparing different methods for fitting voxel-specific TACs is scarce so far. Thus, no generally accepted standard-method ex-

ists. Guidelines recommend either not modeling at all, rather performing trapezoidal integration of TACs, or modeling by exponential functions [159, 145], although these recommendations are usually in the context of whole-organ dosimetry.

In the literature of voxel-based dosimetry in targeted radiotherapy, most groups applied no modeling at all for the time period covered by imaging, and mono-exponential extrapolation after the last imaging [143, 123, 165, 52, 135]. Jackson et al. [69] used tri-exponential modeling of the TAC, but applied regularization in order to limit the number of voxels with implausible TIA. In the study of Kost et al. [82], the user could choose between mono- or bi-exponential modeling of pharmakokinetics. However, the focus of their study was not on the comparison of the different models and consequently they did not report on this.

Marcatili et al. [99] implemented an algorithm which automatically selects the best model for each voxel, either pure mono-exponential decay or linear uptake followed by mono-exponential decay, and no modeling and trapezoidal integration as fall-back if modeling fails. They did not provide a systematic comparison between the different models but stated that the fall-back was used in less than 5% of voxels for their application (I-124 PET). However, it could be hypothesized that for SPECT imaging of typical ^{177}Lu activities, the image noise would be comparably higher than for PET, which potentially could affect fitting of per-voxel TACs negatively.

To the best of our knowledge, the study of Sarrut et al. [136] is the only study that systematically compared different models for voxel-wise TAC fitting. They focused on an algorithm called VoMM, that automatically selected the best model for fitting the TAC of individual voxels. The models considered were mono- or bi-exponential functions with variable degrees-of-freedom of the fit ranging from 2 to 4 parameters. They compared the results obtained by VoMM and by using fixed models for all voxels in a numerical experiment with varying noise levels and also in one dosimetry patient. In any case, 6 imaging time points were used. For the numerical experiment, they evaluated the Root Mean Squared Error between ground-truth and modelled TIAs and also the percentage of voxels which could not be fit successfully ($\%$FF). They found that the Root Mean Squared Error and $\%$FF was lower for VoMM as compared to the fixed models. Notably, they also showed that a low count number and thus high noise level in a voxel's TAC would result in an increased Root Mean Squared Error and $\%$FF. For their patient data, they found that $\%$FF was lower for the VoMM than for most fixed models. However, one fixed model did have an even lower $\%$FF.

Altogether, literature on this topic shows that voxel-wise fitting of TACs, subsequent calculation of TIAs, and thus voxel-specific dosimetry is feasible. However, it also turns out that especially the TAC fitting is most likely negatively affected by noise. Most groups apply some technique to either circumvent mod-

eling of especially the early part of the TACs by using numerical integration or implement some kind of regularization in order to limit the bias introduced by voxel noise. Most groups did only assess if voxel-wise fitting could be done successfully from an algorithmic point of view, but did not evaluate if a successful fit resulted in feasible TIAs. Based on our experience and from our results obtained by the simple fit method, we suspect that many voxels' TIAs, although obtained by successful fitting, are still incorrect. This problem could be alleviated by application of *Particle Filters*, which could be understood as de-noising based on a plausible assumption/model. Altogether, we think that particle filters offer an interesting alternative to methods reported from literature and should be further evaluated in forthcoming studies.

Chapter 9

Dose estimations

This chapter is based on [TIG-3, TIG-4, TIG-5, TIG-6]. The four SPECT-images were integrated over time to obtain the nuclear decays per voxel. Afterwards from this decay map a dose distribution can be calculated. There are different method to calculate the dose. In this chapter seven different methods will be compared to each other.

9.1 Ground truth: Monte Carlo simulations (MC)

In the previous chapter, the number of ^{177}Lu decays per voxel for one therapy cycle was derived from four SPECT images. For calculating patient-specific maps of mass density, the CT part of the SPECT/CT acquisition was used. In CT, the attenuation coefficient μ of tissue for polychromatic X-rays is measured and usually given in *Hounsfield Units* (HU), which expresses the degree of attenuation relative to water and air (see Equation 9.1).

$$H(\mu_{\text{tissue}}) = \frac{\mu_{\text{tissue}} - \mu_{\text{water}}}{\mu_{\text{water}} - \mu_{\text{air}}} \cdot 1000[\text{HU}] \tag{9.1}$$

where $-1024 = -2^{10} \leq H\ [\text{HU}] \leq 4096 = 2^{12}$.

Since attenuation is mainly caused by interactions between photons and electrons, μ is strongly correlated to electron density, which itself is correlated to mass density ρ. For this, a CT image can be used to estimate a mass density map, usually by using a piecewise linear relation, with different slopes for HU ≤ 0 and HU ≥ 0. It can be derived for every CT-scanner by special calibration phantoms. A system-specific CT calibration was not carried out. Instead, mass density was calculated as defined in [8].

$$\rho(x) = \begin{cases} 0.0011 \cdot x + 1 & \text{for } x \leq 0 \\ 0.0007 \cdot x + 1 & \text{for } x > 0 \end{cases} \tag{9.2}$$

where $\rho(x)$ is the mass density in units of $\mathrm{g/cm^3}$ belonging to a voxel with HU number x.

In reality, especially for absorption of beta-particles, not only the mass-density determines absorptive properties but also the varying atomic composition of tissues (e.g. effective mass- and atomic numbers) in every voxel. Unfortunately, an exact determination of these parameters with a single CT acquisition is not possible, but several methods for a rough estimate of tissue type exist. We used a classification based on each voxel's mass density, as proposed in [139]. The assigned tissue-type, composition of tissue-type, and respective range of mass densities can be found in table 9.2 and in table 9.1.

Table 9.1: Prototypical materials ordered by their characteristic atomic composition

Z	A	Atom	1	2	3	4, 5, 6	7	8	9
6	12.01	C	0.0001	0.1023	0.6372	0.2322	0.1078	0.2780	0.1443
7	14.01	N	0.7553	0.0287	0.0080	0.0249	0.0277	0.0270	0.0420
8	16.00	O	0.2318	0.7571	0.2323	0.6302	0.7547	0.4100	0.4461
18	39.95	Ar	0.0128						
1	1.01	H		0.1013	0.1195	0.1045	0.1006	0.0640	0.0472
11	22.99	Na			0.0005	0.0011	0.0008		
15	30.97	P		0.0008	0.0002	0.0013	0.0018	0.0700	0.1050
16	32.06	S		0.0023	0.0007	0.0020	0.0024	0.0020	0.0032
17	35.45	Cl		0.0027	0.0012	0.0013	0.0008		
19	39.10	K		0.0019	0.0003	0.0020	0.0030		
12	24.31	Mg		0.0007	10^{-5}	0.0001	0.0019	0.0020	0.0022
20	40.08	Ca		10^{-5}	10^{-5}	0.0002	10^{-5}	0.1470	0.2099
26	55.85	Fe		0.0004	10^{-5}	10^{-5}	10^{-5}		
30	65.39	Zn		10^{-5}	10^{-5}	10^{-5}	10^{-5}		0.0001

For generating ground truth data, a full MC of radiation transport was carried out for each patient. For this, the GEANT4-based GAMOS 5.0.0 toolkit was used [10]. The procedure is briefly described in the following:

- A voxelized geometry was generated based on the tissue classes and mass-density of each voxel, determined as described before. The size of voxels was $4.8 \times 4.8 \times 4.8\ mm^3$, identical to that of decay, density, and tissue-type maps.

Table 9.2: Prototypical materials ordered by their characteristic densities. The material *Soft-Tissue*, is defined by the atomic composition and aggregated density range of materials marked by *. In the fourth column the percentage ratios of voxels from one tissue class referred to the total number of voxels within the body are illustrated. In the fifth column the percentage dose in each tissue class is shown.

Density range g/cm^3	Material name	Index	ratio of voxel (%)	ratio of dose (10^{-5}%)
$[0.000, 0.207)$	air	1	32.9	7.8
$[0.207, 0.919)$	lung	2	35.6	28.2
$[0.919, 0.979)$	adipose	3	13.5	42.9
$[0.979, 1.004)$	breast*	4	6.3	45.8
$[1.004, 1.043)$	phantom*	5	28.0	69.6
$[1.043, 1.109)$	liver*	6	7.1	58.5
$[1.109, 1.113)$	muscle	7	0.1	38.4
$[1.113, 1.496)$	bone compact	8	3.8	45.9
$[1.496, 1.654]$	bone cortical	9	0.4	44.6

- The decay modes of ^{177}Lu and emitted gamma and beta energy spectra were defined according to [10].
- In total, $1.2 \cdot 10^{10}$ decays of ^{177}Lu were performed using GAMOS on a high-performance cluster (number of nodes: 176).
- The decays per voxel was taken from the decay map, as described before.
- Deposited energy was scored per voxel in units of Gray per decay.

9.1.1 Dose Voxel Kernel (DVK)

For calculating the DVKs, a MC of $2 \cdot 10^8$ of ^{177}Lu decays at a random position in the center voxel of a $9 \times 9 \times 9$ cube of voxels (each $4.8 \times 4.8 \times 4.8\ mm^3$) of certain tissue-type and density was simulated and the deposited energy scored. This was carried out using the GAMOS Monte-Carlo toolkit on a high-performance cluster. The accuracy and precision of results obtained by DVK dosimetry are not only influenced by the method for adapting the DVK to heterogeneous tissue, but also by the statistical uncertainty and size of the DVK. The standard error for individual voxels of the DVK were calculated as described before in section 2.2 , using $M = 10^7$ and $N = 20$. For the analysis of adapting DVKs to different

densities and tissue types, the kernel size was kept constant at $9 \times 9 \times 9$ voxels, each with a voxel size of $4.8 \times 4.8 \times 4.8 mm^3$, which represents a common size in DVK dosimetry [42, 37]. The mean error of the voxel values of a typical Kernel ($t =$ ST, $\rho = 1.04$) with a total number of $2 \cdot 10^8$ simulated decays is $\sim 0.48\%$. The error is smaller for voxels near the center of the DVK, compared to more distant ones, as can be seen in Figure 9.1 for a Kernel with $t =$ ST, $\rho = 1.04$.

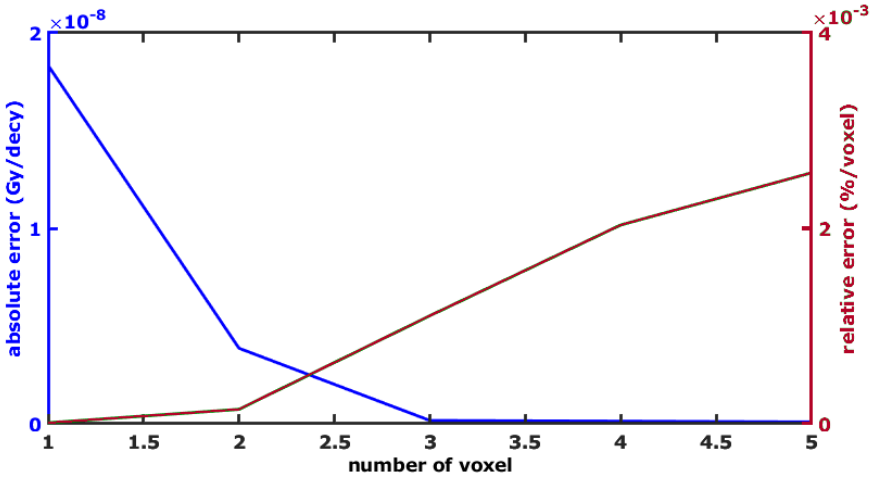


Figure 9.1: Relative error in percent resulting for one voxel from DVK convolution computations.

Having obtained proper DVKs from MC simulations, the error resulting from the DVK convolution with the accumulated activity map needs to be estimated as a next step. Let $k_{i,j,k}$ denote the DVKs for a tissue cube of size $9 \times 9 \times 9 = 729$ voxels. The absolute error for each voxel is designated as $e_{i,j,k}$, and the accumulated radioactivity distribution A is assumed to be uniform. The relative error for the calculated dose value for one voxel $D(\mathbf{r})$ centered at $\mathbf{r}$ can be estimated through the following equation:

$$\frac{D(\mathbf{r}) \pm e_{\mathbf{r}}}{D(\mathbf{r})} = A \cdot \sum_{i,j,k} \left(\frac{k_{ijk} \pm e_{ijk}}{\sum_{i,j,k} k_{ijk}} \right) \tag{9.3}$$

In Figure 6.1, the occurrence of relative errors per voxel, resulting from DVK convolutions and estimated for all 200 computed basic kernels, is illustrated in form of a histogram. The mean error is 0.0113%, which is very small because of the high number of decays which were simulated. The error is of the same size as the error resulting from MC simulations. Consequently, the dose distribution can be estimated with high accuracy.

9.1.2 Statistical error of a whole body simulation

For one representative patient, the error of the MCs was determined (see Figure 9.2). The estimated values of the absorbed dose in a voxel inside the body result in a histogram representing a uniform distribution. The mean absorbed energy dose of this voxel is $D = (0.72 \pm 0.41) \times 10^{-5}\ [Gy/decay]$. If we take into account that 1200 samples were drawn to obtain this mean value, a relative standard error of $rSE = 1.67\%$ results. Altogether, the error of the MC simulations was found to be $rSE \leq 2\%$ for voxels belonging to the evaluated organs. The highest deviation from the mean absorbed dose is obtained in the air surrounding the patient with a maximal error of $rSE = 4.50\%$. Figure 9.2 shows the per-voxel deposited dose and the corresponding rSE estimate at one representative slice position.

Altogether, the achieved precision is similar to comparable studies from literature, e.g. Moghadam et al. [76] state a precision below 2%, but did not disclose how this had been determined. In the study of Dieudonne et al. [37], the authors provide a precision of 4.7% for individual voxels, but did as well not describe the method for determining this number. Overall, the precision of our simulations should be sufficient for ground truth requirements.

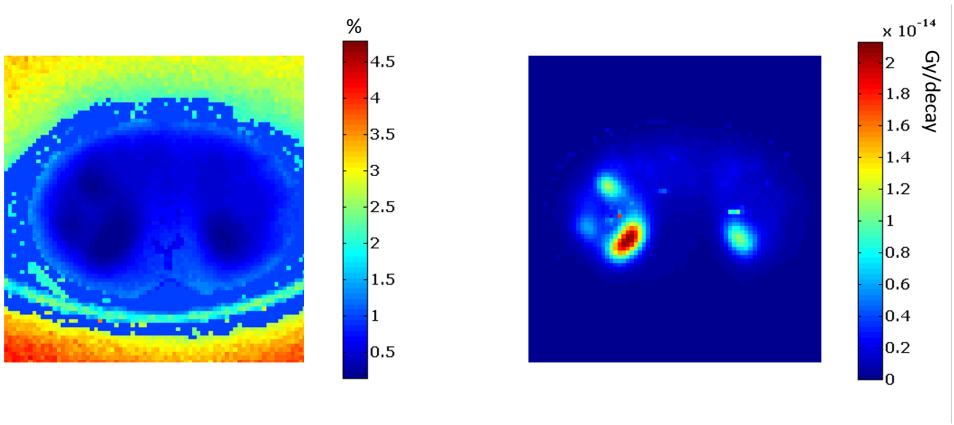


Figure 9.2: On the left side, the result of a MC for one patient is shown. The colorbar reflects an encoding of the deposited dose in $[Gy/decay]$. On the right side, the rSE for the MC is shown. The colors give information about the error in $\%$.

9.1.3 Homogeneous patient

To quantify the error induced by the convolution method for 26 patients, a full Monte Carlo simulation was determined on a homogeneous body consisting of soft tissue with a density of $1\frac{g}{cm^3}$. In addition, a full MC on the CT-densities

was calculated. Both assumed the same distribution of the number of nuclear disintegrations. The total dose per organ was calculated and compared to the dose distribution determined by convolution of the decay distribution with a soft tissue kernel.

The total organ dose for different organs was compared between the soft-convolution method and the full MC_{homo} of a homogeneous patient and the soft-convolution method and the full MC_{CT} on the CT-derived mass density distributions. To characterize differences, the percentage deviation between both methods was calculated for each patient and each segmented organ. The results are collected in the following table 9.3.

Table 9.3: Relative deviation between MC_{homo} und MC_{CT} in percent.

organ $\diagdown$ method	MC_{homo} vs. STS	MC_{CT} vs. STS
left kidney	$5.67 \pm 1.96\%$	$8.62 \pm 8.21\%$
right kidney	$6.11 \pm 3.28\%$	$13.81 \pm 17.77\%$
spleen	$7.45 \pm 8.13\%$	$13.37 \pm 16.92\%$

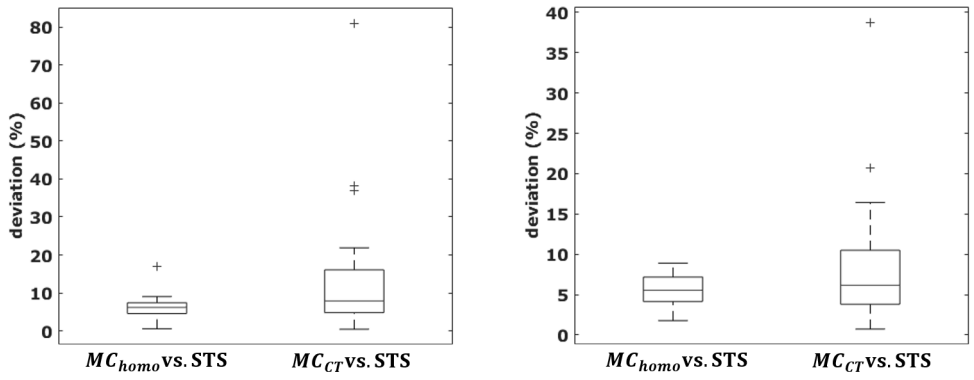


Figure 9.3: Percentage deviation of total dose values of the left kidney (left side) and right kidney(right side) between three methods for 24 patients.

Note that any estimation, with the neural network, of the dose voxel kernel (DVK) based on heterogeneous mass density kernels, cannot estimate the whole organ dose better than 5.67% for the left kidney, 6.11% for the right kidney and 7.45% for the spleen. Relative to the current results, any improvement must achieve values better than 8.62%, 13.82% and 13.37%. With the large patient

cohort, we know for each patient the limiting values and can precisely evaluate the quality of our neural network results.

To evaluate the size of the DVK, the MC for one patient was determined on an artificial body only consisting of soft tissue instead of the CT image. Therefore, only the influence of the activity distribution and the DVK size can be seen in this simulation. With increasing size of the DVK used for convolving the decay distribution, the mean deviation relative to values from full MC simulations decrease, as can be seen in Table 9.4.

Table 9.4: Deviation (in $\%$) for dose values obtained by DVK-Dosimetry, relative to full MC simulations in dependence of the used kernel size.

DVK-Size (Voxels)	left kidney ($\%$)	right kidney ($\%$)	Spleen ($\%$)
81^3	2.00	3.80	1.68
51^3	2.15	3.90	1.79
31^3	2.32	4.09	1.98
9^3	3.81	5.72	3.56
7^3	4.15	6.11	3.90
5^3	4.54	6.53	4.28
3^3	4.95	6.96	4.68

In our evaluation of the influence of DVK size, it was shown for dose determination in soft-tissue that DVK methods systematically underestimate deposited dose, compared to values obtained from full MC simulations. Most likely, this is due to the truncation of radiation transport by the limited size of a DVK, whereas radiation transport in the MC simulations is only limited by the comparably larger size of the simulation volume. Consequently, this underestimation decreases with increasing size of the DVK.

9.2 Dose estimation with DVK via Scaling

In this chapter, the four methods *Soft Tissue Scaling* (STS), *Center Scaling* (CS), *Density Scaling* (DS) and *Percentage Scaling* (PS) will be compared with each other. As ground truth, the full MC was used. The organ and tissue type-wise dose estimation is evaluated.

The following considerations refer to dose kernels $K(\mathbf{r}_T \leftarrow \mathbf{r}_S)$ consisting of $9 \times 9 \times 9$ voxels, where each voxel represents a volume of $V_{vox} = 4.7^3\ [mm^3]$.

9.2.1 Soft Tissue Scaling (STS)

The simplest way to obtain a dose distribution from the decay map is to convolve the decay map with a DVK corresponding to soft tissue with a mass density of $1.0\ \left[\frac{kg}{dm^3}\right]$. This standard method, simply called soft scaling henceforth, assumes a homogeneous patient consisting only of soft tissue. Dose inhomogeneities because of density changes are not taken into account, rather only inhomogeneities because of spatially varying activity distributions are accounted for. Henceforth such DVKs will be called *soft tissue kernels* $K^{st}_{vox}(\mathbf{r}_T \leftarrow \mathbf{r}_S)$ for short.

9.2.2 Center Scaling (CS)

If the dose kernel consists of several voxels of possibly differing tissues, almost 60% of the energy relieved during a nuclear disintegration is deposited in the center voxel. For this reason, one could come to the conclusion that the soft tissue convolution can be improved by convolving the mass density distribution not only with one soft tissue kernel but rather with 200 tissue kernels $K^{m}_{vox}(\mathbf{r}_T \leftarrow \mathbf{r}_S, \rho_m)$ corresponding to tissue mass densities ρ_m between $10\ \left[\frac{kg}{m^3}\right]$ and $2000\ \left[\frac{kg}{m^3}\right]$. To each voxel of the decay map a DVK is assigned which corresponds to a tissue density which is closest to the mass density of the respective CT-voxel.

9.2.3 Density Scaling (DS)

As mentioned above, the result of each MC simulation is a DVK $K_{vox}(\mathbf{r}_T \leftarrow \mathbf{r}_S)$ in the unit $Gy/decay$. It is related to the mean energy imparted to matter inside a voxel per unit mass by ionizing radiation resulting from a single nuclear disintegration $a(\mathbf{r}_S, t)$:

$$K_{vox}(\mathbf{r}_T \leftarrow \mathbf{r}_S) = \frac{D_E(\mathbf{r}_T)}{A(\mathbf{r}_S)} \tag{9.4}$$

$$D_E(\mathbf{r}_T) = \frac{\epsilon(\mathbf{r}_T)}{\rho_m(\mathbf{r})} \tag{9.5}$$

where $\epsilon(\mathbf{r}_T)$ represents the energy density and $\rho_m(\mathbf{r})$ the spatially varying mass density inside the voxel. Consequently, the absorbed energy dose depends strongly on the mass density of the tissue inside the voxel. A simple work-around the plain soft tissue approximation was proposed in [37] by considering the absorbed energy density $\epsilon(\mathbf{r}_T)$ as a system invariant, i. e.

$$D_m(\mathbf{r}_T) \cdot \rho_m(\mathbf{r}_T) \equiv \epsilon_m(\mathbf{r}_T) \overset{!}{=} \epsilon_{st}(\mathbf{r}_T) \equiv D_{st}(\mathbf{r}_T) \cdot \rho_{st}(\mathbf{r}_{st}). \tag{9.6}$$

Here $D_{st}(\mathbf{r}_T)$ denotes the DVK, which is calculated in a MC for a target voxel containing soft tissue with mass density $\rho_{st} = 1040\, kg/m^3$. Furthermore, $\rho_m(\mathbf{r}_T)$ represents the voxel mass density of the actual target tissue considered, and $D_m(\mathbf{r}_T)$ represents the corresponding density corrected dose value. Hereby, it is assumed that the deposited energy density $\epsilon_m \approx \epsilon_{st}$ is largely independent from the tissue composition. This assumption might be reasonable if the atomic composition of the considered tissues is roughly similar to soft tissue. However, it might be expected to fail in case of materials like bones or lung. The idea behind this first order approximation is that the elementary radiation - matter interactions stay the same, only the event probability changes with density.

9.2.4 Percentage scaling (PS)

Rather than taking into account different tissue densities for DVK calculations via a density scaling method as discussed above, in this thesis a new method is proposed to account for tissue heterogeneity following ideas first expressed in [76].

Consider a kernel whose volume encompasses N_V^3 voxels, where the voxel size plays only a minor role as long as its length dimension is larger than the effective range of the emitted radiation. A radioactive isotope is placed in the center voxel located at $\mathbf{r}_T \in \mathbb{N}^3$ on a regular grid, where the nodes are index with (i, j, k). In the following, the DVK $S(i, j, k)$ will be computed based on a heterogeneous mass density distribution discretized to the length scale of a voxel (typically $5\ [mm]$. To do so, the DVKs $S_m(i, j, k)$ for all tissues with mass density ρ_m which make up the kernel volume $V_K = N_V^3 \cdot V_{vox}$ are needed. The dose values for such a heterogeneous kernel at the target location $\mathbf{r}_T = (r_i r_j r_k)^T$ can be estimated as follows ($r_i \equiv i, r_j \equiv j, r_k \equiv k$ for short):

$$
\begin{aligned}
D_m(i,j,k) \quad &= \quad \frac{D_{st}(i,j,k)}{D_{st}(c,d,e)} \cdot D_m(c,d,e) \\
&\text{with} \\
c = i+1, d = j+1, e = k+1 \quad &\text{if} \quad c < 0, d < 0, e < 0) \\
c = i-1, d = j-1, e = k-1 \quad &\text{if} \quad c > 0, d > 0, e > 0)
\end{aligned}
$$

where the dose delivered to any voxel of the kernel depends on the dose deposited in the neighboring voxels which are located closer to the center voxel. The percentage decrease of the dose in the homogeneous kernel is used to scale the dose value in the heterogeneous kernel. We will refer to this method as the 'Percentage'-method in the following.

9.3 Dose estimation with DVK via NN (DVK-NN)

9.3.1 Network architecture

The network architecture for a 3D convolutional U-Net, as employed in this study, is illustrated in figure 9.4. The architecture of the U-Net depends on the size of the input data array, here the mass density kernels. We choose 3D density kernels consisting of $9 \times 9 \times 9$ voxels corresponding to a physical volume of roughly $45\ mm^3$.

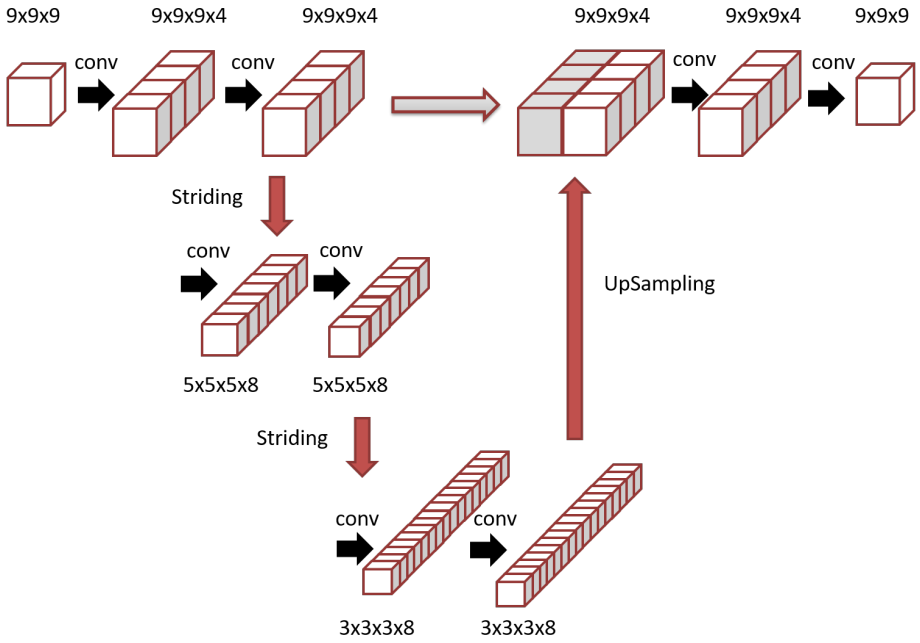


Figure 9.4: Network architecture, where each convolutional layer includes a leaky ReLu activation layer and a batch normalization layer.

Each convolutional filtering operation is followed by an activation using a leaky ReLU activation function. Furthermore, batch normalization is applied also. Convolutional filtering was performed with a $3D$ filter size of $3 \times 3 \times 3$ voxels and was always accompanied by padding to preserve the size of the input array.

The following *convolutional filtering* possibilities emerge:

- The first convolutional step will employ four filters to create four feature maps
- The second convolutional step uses only one filter for each feature map. This keeps the number of feature maps constant but increases the receptive field size in the input data array.

Downsampling was done employing striding and pooling. Starting with a $9 \times 9 \times 9$ density kernel and applying a stride of one and average pooling with a $3 \times 3 \times 3$ pooling filter, we get the following sequence of feature kernels :

$$9 \times 9 \times 9 \rightarrow 5 \times 5 \times 5 \rightarrow 3 \times 3 \times 3 \rightarrow 1 \times 1 \times 1$$

The number of filters has been doubled at each down-sampling step. Here the pooling filter simply slides along the Cartesian axes with a step size of one voxel and condenses context information into the voxels of the resulting feature map while keeping the spatial location relative to the input data array roughly constant. This results in at most 3 down-sampling steps with average pooling but with no voxel information being skipped. The last down-sampling step does not seem to be useful and has been omitted.

$$9 \times 9 \times 9 \rightarrow 5 \times 5 \times 5 \rightarrow 3 \times 3 \times 3$$

Upsampling is needed to arrive at an output array with the same dimensions as the input array. Starting with the $3 \times 3 \times 3$ array, upscaling has been performed by linear nearest neighbor interpolation to yield a $9 \times 9 \times 9$ array.

For training, the following loss function was used:

$$loss = \sum_{i}^{729} \left| \log\left(\frac{a_i}{\hat{a}_i}\right) \right| \cdot \frac{\log(a_i)}{\sum_j^{729} \log(a_j)} \tag{9.7}$$

where $\log(\hat{a}_i)$ and $\log(a_i)$ are the logarithms of the dose values $\hat{a}_i$ estimated by the neural network and a_i the dose values calculated by a Monte Carlo simulation, respectively. Logarithmically scaled dose values were chosen because of numerical problems with representing the original dose values.

For each voxel a_i of the DVK determined through a MC, the error Δa_i is known. Therefore, the value of the loss function is limited by the precision of the MC.

Noting that $\hat{a}_i = a_i \pm \Delta a_i$, hence the loss function can be re-written as follows:

$$\begin{aligned} loss &= \sum_{i}^{729} |\log(a_i) - \log(\hat{a}_i)| \cdot \frac{10 - \log(a_i)}{\sum_j^{729} \log(a_j)} \\ &= \sum_{i}^{729} |\log(a_i) - \log(a_i \pm \Delta a_i)| \cdot \frac{10 - \log(a_i)}{\sum_j^{729} \log(a_j)} \\ &= \sum_{i}^{729} \left| \log\left(\frac{a_i}{a_i \pm \Delta a_i}\right) \right| \cdot \frac{10 - \log(a_i)}{\sum_j^{729} \log(a_j)} \end{aligned} \tag{9.8}$$

With 20785 training kernels an average loss value of $6.4 \cdot 10^{-5}$ resulted.

The loss function takes into account that it is important to have a precise dose estimation for the center voxel with a generally high dose value rather than for the peripheral voxels with small dose values.

Through parameter evaluation, the first loss function (eqn. 9.7) yielded much better results than the second one. Also the hyperparameters α of the leaky ReLU-function, the batch size N_{mb} and the number of filters N_{conv} were evaluated. The best results were achieved, if the hyperparameters were set as given in table 9.5.

Table 9.5: Hyperparameters of the loss function

number of filters in the first convolution layer	$N_{conv} = 4$
leakage parameter of the ReLU-function	$\alpha = 0.1$
mini-batch size	$N_{mb} = 32$

9.3.2 Training data and evaluation

This study focuses only on the prediction of dose values for the kidneys. Density kernels with a size of $9 \times 9 \times 9$ voxels were taken from the extracted organ volume (see chapter 6.3) with a shift of one voxel. Consequently, for each patient and each kidney we end up with a set of density kernels which sum up to one entire kidney.

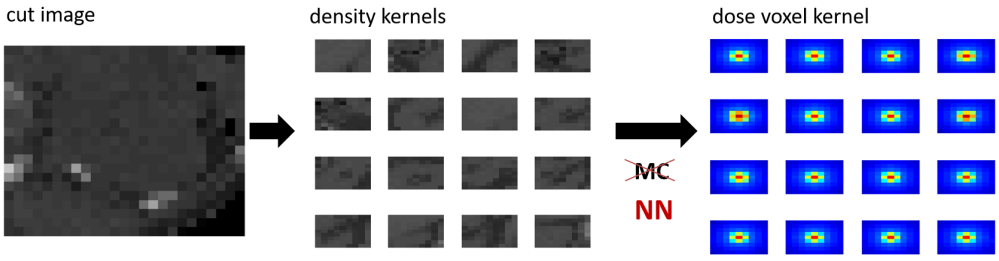


Figure 9.5: On the left side, a CT image from the left kidney, divided in density kernels, one of which is illustrated in the middle. With a MC, the DVK has been calculated (seen on the right side). The MC should be replaced by a NN

Four of the 26 patients were chosen randomly for training the network. These 8 kidneys consisted of 31000 density kernels encompassing $9 \times 9 \times 9$ voxels. For each density kernel, a Monte Carlo was performed with $2 \cdot 10^8$ decays. The MC parameters were chosen identically to those used in the calculation of DVKs as explained in chapter 9.1.1. Next, these mass density kernels were used, together

with the corresponding DVKs as obtained from the MC simulations, to train the NN (see figure 9.5).

After training, the kidneys of the remaining 22 patients were predicted. The average dose values for the kidneys, as predicted by the NN, were compared to the corresponding dose values as resulting from full MC simulations. The NN estimates DVKs. With each estimated $9 \times 9 \times 9$ DVK, the dose absorbed in one voxel can be calculated by way of a convolution operation. Now, the DVK of the center voxel is the most important, because on average 70% of the dose is absorbed there. Consequently, if the DVK of the center voxel is estimated very badly, the dose value will be far away from the truth. For this reason, the DVKs of most of the voxels of a kidney can be estimated very precisely but the volume can also contain strong outliers. To get rid of these outliers, a statistical data analysis was applied after the estimation of the dose values from any kidney.

To detect outliers in the estimated dose distribution of a kidney, we use the modified Thompson τ technique for outlier detection [7]. It is defined as:

$$\tau = \frac{t_{\frac{\alpha}{2}} \cdot (n-1)}{\sqrt{n} \cdot \sqrt{n-2+t_{\frac{\alpha}{2}}}}$$

where n is the number of data points, $t_{\frac{\alpha}{2}}$ is the critical student's t value, based on $\alpha = 0.05$.

After deleting the outliers, the mean dose values for the kidneys have been calculated.

However, after training the estimation of the kidneys of the 22 patients were not yet satisfactory. Therefore the training data set was extended. But as the MC simulation of the density kernels is very time consuming, it was not possible to simulate all density kernels from a larger number of kidneys. Therefore, we decided to select four additional patients, where the deviations of the predictions from the MC results were the largest. From these patients, we extracted for each kidney all voxels with a larger than 3% deviation from the MC result. These voxels then entered in an additional MC. The new DVKs were then added to the training data set and the NN was trained again. In summary, finally the kidneys of 8 patients with altogether 21274 density kernels were used.

The new training was performed with 52274 data sets which each was divided in 80% training data and 20% validation data. The loss functions for training and validation in dependence on the number of epochs is illustrated in Fig. 9.6.

With the trained NN, the DVKs of 28 kidneys from 14 patients were predicted. Afterwards the predicted DVKs of each kidney were reconstructed to obtain the voxel-wise dose values for the entire kidney. Any outliers were deleted beforehand and the mean dose per kidney was calculated and is denoted in the following as *DVK via Neural Networks* (DVK-NN).

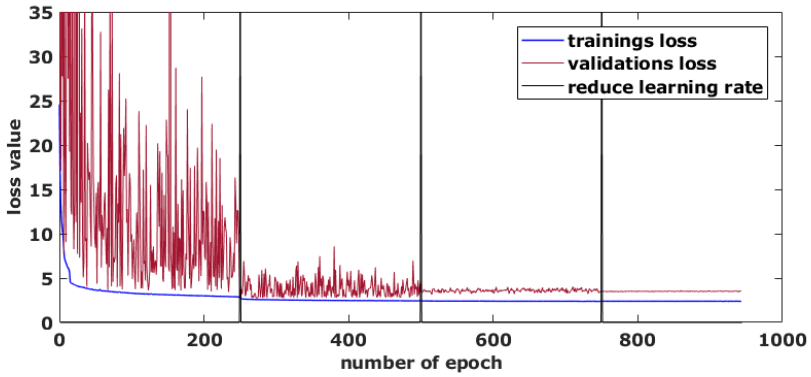


Figure 9.6: Training and validation loss function in dependence of the number of epochs.

The center voxel of every DVK is the most important voxel as it contains the radiation source. In Fig. 9.7 the percentage deviations between the voxel dose values calculated by the MC simulation compared with the NN method are illustrated in dependence on the density of the center voxel. It is shown that the densities around $\rho_m = 1\frac{g}{cm^3}$ can be estimated very precisely with the NN method. The reason for this is that the training data which was used is not uniformly distributed across the given density range as can be seen in Fig. 9.7, where the occurrence of each density within the training data is presented. Most training data consists of density kernels with a center voxel density close to $1\frac{g}{cm^3}$.

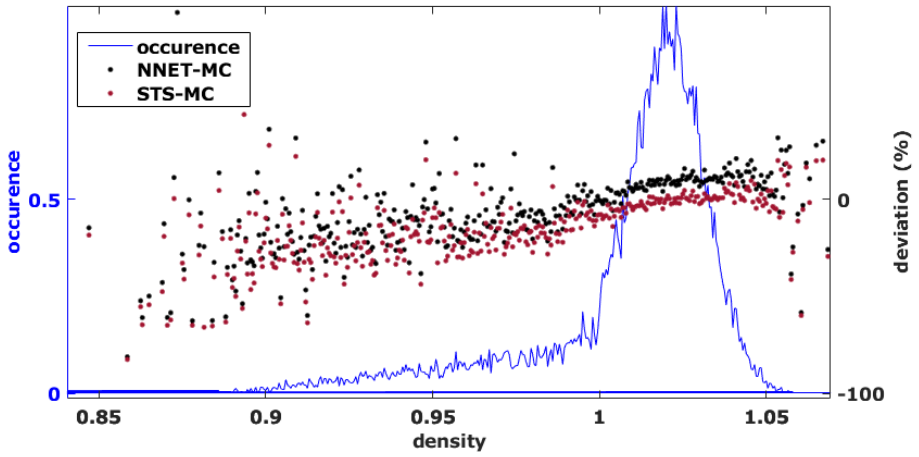


Figure 9.7: Percentage deviation between MC and the two methods in dependence on the density of the center voxel of the density kernel. The occurrence of the density within all kidneys can also be seen on the ordinate.

9.4 Dose estimation of whole patients via NN

9.4.1 Preprocession

The body of each patient was segmented from the CT image and the SPECT-related decay map to reduce computation time. Therefore, the data sets have different sizes in each spatial dimension. Along the x- and y-directions, the images were rescaled to a size of 80×80 pixels.

The data sets of the 26 patients had different sizes. For better performance of the DNN, 5 adjacent slices were selected from the 3D imaging data (CT, SPECT) of each patient. These imaging data from all patients were rescaled to an 80×80 voxel array in $x-$ and $y-$ direction (see figure 9.8). Taking only 5 slices is motivated by the physical nature of the emitted radiation of Lutetium, which is a β^- emitter. The range of the emitted β-radiation is at most two voxels wide, given a voxel size of $(4.89\ mm)^3$. Therefore, to estimate the dose of the center slice it would be enough to know the decay distribution and the density distribution of two slices neighboring the center slice along the z-direction. So the idea is to have two $3D$ inputs to a DNN and one $2D$ output in a U-Net-like DNN architecture. All images were normalized to the respective maximal value of the entire data set.

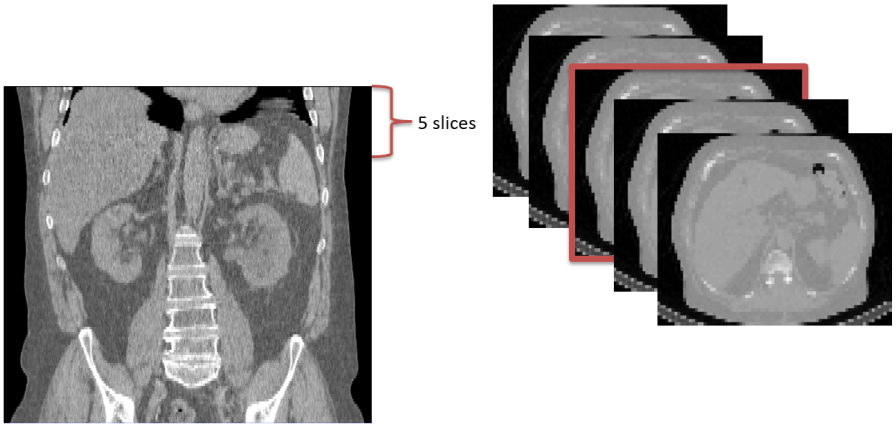


Figure 9.8: Selection of 5 slices from one patient used for estimating the dose of the center slice.

9.4.2 Net architecture

The study intended in letting a DNN learn, in a supervised fashion, the true dose map as it resulted from a full MC of the whole body of the patient. As input, the mass density map and the dose map, as calculated with the simple STS method, were provided. To achieve this goal, and to evaluate the dose map as predicted by the DNN, for each patient the kidneys, the spleen and all tracer positive metastases had to be segmented by a medical expert first.

In figure 9.9, the architecture of the DNN is illustrated. One input is given by the density map and the other input corresponds to the approximate dose map, as calculated with the standard method. Each input data array had a size of $(80 \times 80 \times 5)$ voxels. Two consecutive convolutions with appropriate filters, followed by a mini-batch normalization and a leaky ReLU activation function, were applied. Afterwards, the data tensor was down-sampled by max-pooling. These steps were repeated four times until the tensor had a size of (5×5) voxels. Next, the two down-sampled input tensors were concatenated and convolved as before. The combined matrix was then up-sampled and convolved until it reached its final size of (80×80) pixels.

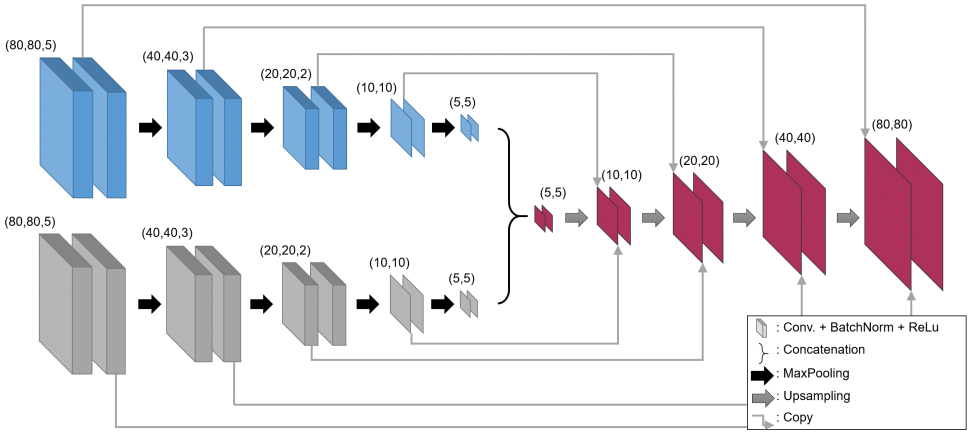


Figure 9.9: Net architecture with two 3D inputs and one 2D output.

The mini-batch size had to be set to one because of limitations of the main on-board memory. The filter size was varied from (5×5), (7×7) and (9×9), and also the number of filters was varied from 4 to 8 and finally to 16 filter maps. As loss function, the mean squared reconstruction error of the entire dose map was used according to

$$L_{wSE} = \frac{1}{N_{voxel}} \sum_{\mathbf{r} \in \Omega} \left(D(\mathbf{r}) - \hat{D}(\mathbf{r}) \right)^2 \tag{9.9}$$

where N_{voxel} denotes the number of voxels in the dose map at the output of the net. Also $D(\mathbf{r})$ denotes the dose map as obtained from a whole body MC simulation, while $\hat{D}(\mathbf{r})$ denotes the corresponding dose map as obtained from the hybrid NNET.

Based on the idea that a two-dimensional GiT-BEMD provides intrinsic modes, which reflect the characteristic spatial scales in their textures of the given mass density and activity maps, the DNN was modified by replacing the CT input by one out of six IMF obtained from a GiT-BEMD. The six IMFs for the slice in figure 6.2 are illustrated in figure 9.10.

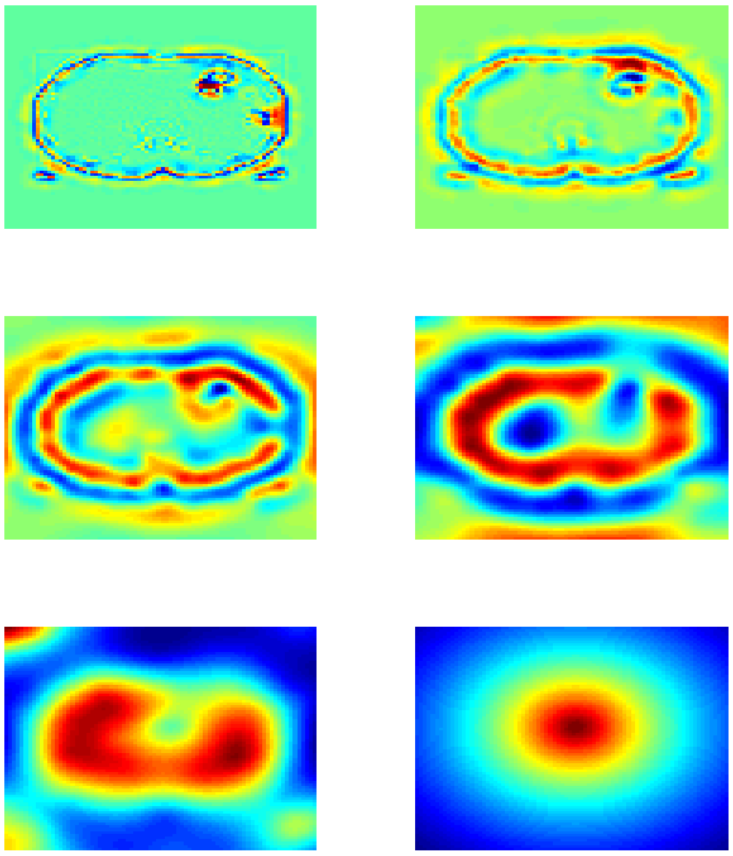


Figure 9.10: The six IMFs for a patient suffering from a prostate cancer.

Parameter evaluation

Parameter evaluation was performed as follows: First from the 3D X-ray CT images of all 26 patients, 1385 packs of 5 adjacent density map slices were extracted, thus yielding a total of $5 \cdot 1385$ slices. This dataset was divided into a training and validation dataset encompassing $5 \cdot 1035$ density map slices and a hold-out dataset corresponding to $5 \cdot 350$ density map slices from the remaining 6 patients used for testing. Note that each input dataset, for instance each density map, or their equivalent IMs, or activity map, encompassed 5 adjacent slices always. The training set contained $5 \cdot 779$ density map slices from 16 patients and the validation dataset amounted to 256 density map slices from the remaining 4 patients. The latter was used repeatedly during training to track the generalizability of the learning process and to avoid overfitting. Training was terminated whenever the training error did not change for the last 10 epochs within a margin of $\epsilon \leq 10^{-5}$ or the validation loss started to increase. Finally the hold-out dataset was used to estimate the prediction accuracy and get an estimate of the variance of the parameter estimation. For better statistics 10 bootstrap samples were drawn from the training and validation datasets resulting in a total of 7790 density maps for training and 2560 datasets for validation, respectively. The resulting MSE of the test dataset was compared for different parameter configurations. Finally, the resulting best parameter configuration was chosen to perform a leave-one-out cross-validation (LOOCV) with all 26 patients. This helped to estimate, in a clinical context, the accuracy with which whole organ doses could be estimated with the newly proposed method.

Without EEMD The validation losses for the different settings are collected in the following table 9.6:

Table 9.6: Testing losses are given in units of $10^{-2}Gy^2$ for different hyperparameter settings. FS denotes filter size and NF counts the number of filters used. The stars mark a significant difference between a configuration and the best one ($p < 0.05$).

FS \ NF	4	8	16
$(5 \times 5 \times 5)$	1.4 ± 1.3	1.5 ± 1.2	$2.9 \pm 3.5 *$
$(7 \times 7 \times 5)$	3.6 ± 2.3	$\mathbf{1.0 \pm 1.2}$	1.4 ± 2.1
$(9 \times 9 \times 5)$	1.5 ± 2.2	$2.0 \pm 2.7 *$	3.4 ± 2.0

The testing loss $L = 1.0 \cdot 10^{-2}\, Gy^2 = (1.0 \cdot 10^{-1})^2\, Gy^2$ achieved lowest values for a configuration with eight filters and a filter size of 7×7 pixels. The differences between most estimates are not statistically significant according to a Wilcoxon sign rank test. Only two estimations are significantly different, with a p-value of

$p < 0.05$, from the selected configuration with smallest test loss. These significant losses are marked with a star in the table. The missing statistical significance of most other values is a consequence of the small number of patients in the cohort. Still the minimal loss configuration was used for all subsequent calculations.

With EEMD Based on the idea that an EMD provides intrinsic modes (IMs), which reflect characteristic spatial-frequency scales and related textures of the given density and activity maps, the U-net was modified by replacing the density maps $\mathbf{M}(\mathbf{r})$ by their corresponding sets of IMs resulting from a slicewise Greens-function-in-tension-based bidimensional ensemble empirical mode decomposition (GiT-BEEMD) analysis [5], where a thorough discussion of the impact of the different model parameters is given. Finally, a principal component analysis (PCA) has been applied to reduce the dimensionality of the IMs. The IMs obtained from decomposing the density map slice in Fig. 6.2 are illustrated in Fig. 9.10. The idea behind using a GiT-EEMD and the related IMs is the following: CNNs extract feature maps which are composed of textures on spatial scales depending on the chosen filter size. But this is what an EMD also achieves. However its related IMs, which are characteristic of the inherent textures contained in the input stimulus patterns, reflect the statistical properties of the intensity distributions of the stimuli. Hence it might be profitable to present partial density maps, which only contain inherent textures on characteristic spatial scales instead of complete density maps with a mix of all inherent textures. This might ease the learning process and help to reduce the variance of the output data arrays.

Consequently, the density maps were replaced either by one of their IMs or by certain combinations of them, as extracted individually from all patients. Again, the optimal network configuration, as deduced above, was used here. The combination of three IMs achieved lowest training $(0.7 \cdot 10^{-2} Gy^2)$ and validation $(0.9 \cdot 10^{-2} Gy^2)$ losses, which with respect to the resulting standard deviations were even superior to the one obtained with the density map as input. Therefore this configuration was used to evaluate the network as described in the following section.

Network validation

The data set is taken from 26 patients and encompasses 27 up to 96 slices along the z-direction, depending on the field of interest. For the validation of the neural network, a *Leave-One-Out Cross-Validation* (LOOCV) was used. The neural network was trained 26 times, whereby each time one patient was left out. Afterwards, the left-over slices were estimated by the NN. For all patients, the deviation between the mean dose within the VOIs, as estimated by the NN or by the Monte

Carlo simulation, was calculated and denoted in the following as *Dose Estimation via Neural Network* (DE-NN). The LOOCV results of the NN with EMD are called *Dose Estimation via Neural Network and EMD* (DE-NN-EMD).

9.5 Comparison of different methods

9.5.1 Results

To characterize the dissimilarities between two dose distributions, the relative organ-wise deviation $\Delta_X(\text{Organ})$ for estimation method X is defined as:

$$\Delta_X(\text{Organ}) = \frac{\sum_{n=1}^{N_{\text{Organ}}}[d_{\text{MC},n} - d_{X,n}]}{\sum_{n=1}^{N_{\text{Organ}}} d_{\text{MC},n}} \tag{9.10}$$

with the dose value $d_{X,n}$ of voxel n calculated using DVK-method X, and $d_{\text{MC},n}$ the dose value of the same voxel computed by the full Monte-Carlo simulation. N_{Organ} is the total number of voxels in the organ and X denotes either STS, CS, DS, PS, DVK-NN, DE-NN or DE-NN-EMD.

In Table 9.7, the average organ-wise deviations of DVK dosimetry relative to full MCs dosimetry can be found. Altogether, the DVK methods underestimate the dose by about $\sim 8\ to\ 10\%$.

Table 9.7: Organ-specific, averaged absolute percentage deviations between dosimetry based on full Monte-Carlo simulation and on DVK-convolution using different method for adapting kernels to tissue inhomogeneities. The numbers indicate average values over the population of 26 patients ± one standard-deviation

	mean (%)	std (%)	max (%)	min(%)
left Kidney				
STS	−7.31	13.85	−42.37	10.22
PS	−8.79	13.63	−43.41	8.40
CS	−8.74	13.63	−43.40	8.44
DS	−9.23	13.60	−43.46	7.97
DVK-NN (only 14 kidneys)	−6.69	10.00	−26.00	7.93
DE-NN	−3.33	5.83	−13.43	4.76
DE-NN-EMD	−3.65	4.37	−12.21	2.13
right Kidney				
STS	−8.48	16.18	−44.76	15.13
PS	−10.03	15.87	−45.95	13.31
CS	−9.92	15.79	−45.93	13.36
DS	−10.30	15.94	−46.15	12.74
DVK-NN (only 14 kidneys)	−6.74	12.77	−26.11	11.02
DE-NN	−3.27	5.69	−13.49	4.02
DE-NN-EMD	−3.75	4.25	−12.18	2.13
Spleen				
STS	−7.27	16.61	−41.12	20.97
PS	−9.73	16.17	−42.71	17.78
CS	−9.66	16.18	−42.67	17.81
DS	−8.06	16.52	−41.55	19.88
DE-NN	−3.58	6.01	−20.28	4.54
DE-NN-EMD	−3.67	4.67	−12.40	4.57
Tumor				
DE-NN	−5.60	6.42	−20.28	4.54
DE-NN-EMD	−4.62	4.78	−12.75	3.43

The percentage differences in dose for the right kidneys of all seven methods are illustrated in figure 9.11.

The tissue-specific relative deviation $\Delta_X(\text{Tissue})$ for estimation method X is defined analogously to the organ-specific deviation $\Delta_X(\text{Organ})$.

In Figure 9.12, the average deviations of estimation dosimetry relative to full

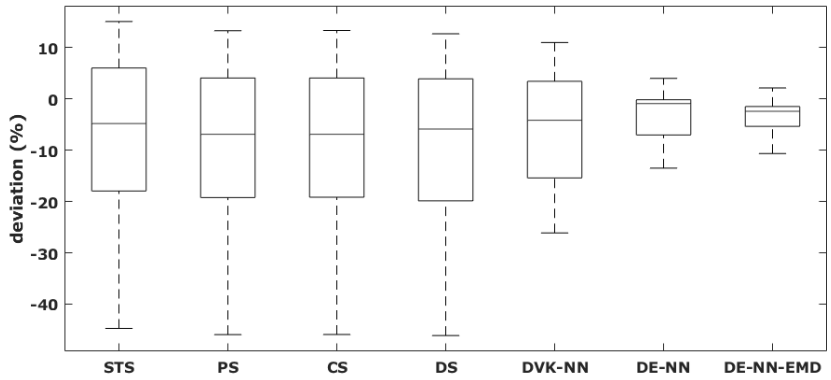


Figure 9.11: Differences in dose for right kidneys compared between seven methods.

MC simulations dosimetry can be found, separately for lung tissue and cortical bone. Two patients were excluded from the analysis for class 9 (cortical bone), since large metallic implants were present in the CT images, which led to extended image artifacts and prevented a meaningful determination of density information.

9.5.2 Discussion

The presented investigation studied the use of several methods from literature for adapting DVKs to tissue inhomogeneities and neuronal network approaches, in order to increase dosimetry accuracy without the need to perform elaborate and time-consuming MCs for each voxel size, density, tissue-type, and radionuclide in clinical use. In this study, 26 ^{177}Lu therapy patients from clinical routine are included, for whom the results from dosimetry and full MCs dosimetry were compared for several organs and tissue types. The MCs were performed employing the GAMOS toolkit, which subsequently acted as ground truth with respect to the absorbed energy dose in each voxel.

When analyzing the results of the organ-specific dose deviations between the used methods and full MCs, it can be seen that on average the presented DVK-scaling methods obtained similar results, while the neural network approaches yielded much better results. Generally, an underestimation of dose by DVK-convolution is seen. On average, the dose is underestimated by about ~ 5 to 8%.

This is similar to the results from other groups which compared DVK- and MC simulations-dosimetry. For example, Dieudonne et al. report an average underestimation of between 2.2% and 5.6% for organs liver, spleen, and renal pelvis medulla. Unfortunately, their results were based on one ^{177}Lu patient only. In a simulation study based on the Zubal phantom, Moghadam et al. [76] found

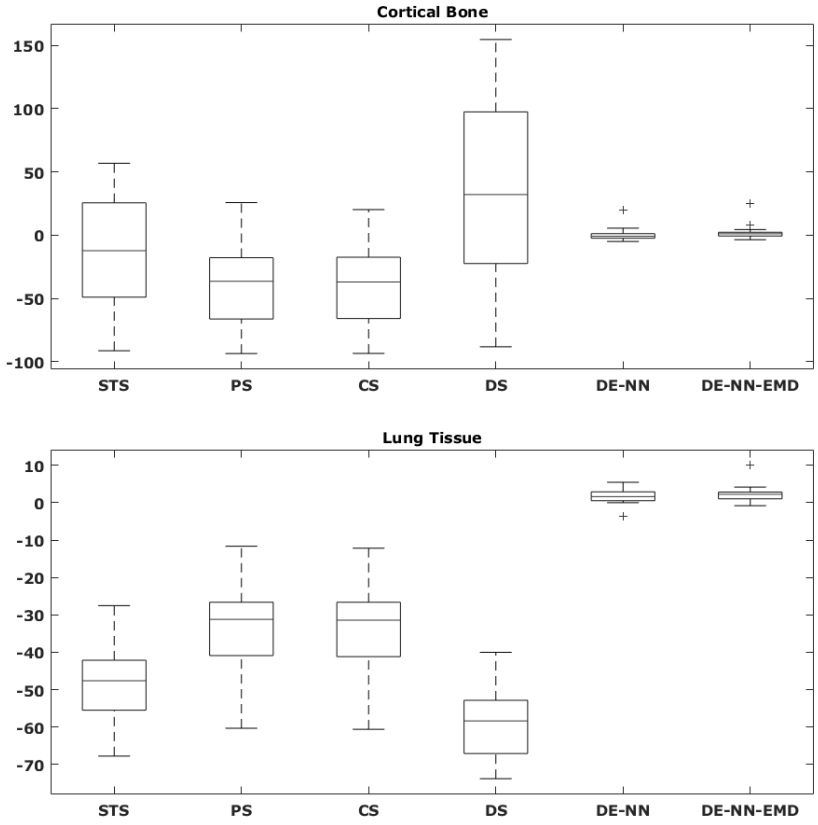


Figure 9.12: Percentage deviations ($\%$) for four different methods ordered according to the tissue class lung tissue, and cortical bone.

an average deviation of 5.2% between MC simulation and DVK dosimetry for tumorous and normal tissue. The trend of underestimation, seen in our study and by others, could most likely be explained by the finite size of DVKs, as has been shown in our experiment using different kernel sizes (Section 9.1.3). In fact, the deviation from the average organ dose estimated by DE-NN and DE-NN-EMD is in the range of the statistical error of the MC. Also the distributions of dose estimates in every voxel agree very well with the corresponding MC results.

Since other tissue classes besides previously discussed organs could be of interest for ^{177}Lu dosimetry, e.g. in patients with lung, lymph node, or bone metastases, we also evaluated the deviations between results for dosimetry methods and full MCs for other tissues, namely lung, adipose, muscle and cortical bone tissue. Especially for the low- and high density tissue classes, compared to soft tissue, such as lung tissue or cortical bone, DE-NN and DE-NN-EMD achieve the best results. The other four methods, STS, PS, CS and DS, represent bad approxima-

tions if the tissue density differs strongly from soft tissue,

In the other studys, like [76] or [37], any scaling algorithm was evaluated only on an artificial phantom or one patient. These study, however, were strongly focused on soft-tissue organs. Generally, the errors increase as the tissue properties and densities differ more strongly from soft-tissue. Altogether, when looking at the results for different types of tissues, the DE-NN-EMD performed best.

9.6 Organ dose calculation via deep neural networks

9.6.1 Data statistics

For each patient, the dose values were calculated voxel-wise. For each VOI, a histogram of all dose values has been established, and an example of such a histogram is illustrated in figure 9.13. The shown distribution is asymmetric and heavy-tailed. Such distributions can be approximated by alpha-stable distributions [51, 131].

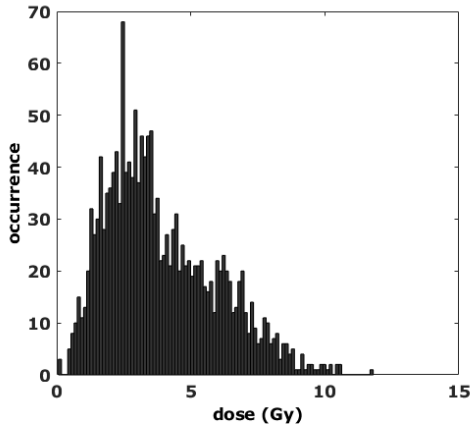


Figure 9.13: Histogram of all dose values for a PSMA-positive bone metastasis.

An alpha-stable distribution, which is characterized by four moments, can be described by

$$\phi(\omega) = \begin{cases} e^{-|\gamma\omega|^{\alpha}[1-i\ sign(\omega)\beta tan(\frac{\pi\alpha}{2})]+i\mu\omega}, & \text{for } \alpha \neq 1 \\ e^{-|\gamma\omega|[1+i\ sign(\omega)\frac{2}{\pi}\beta log(|\omega|)]+i\mu\omega}, & \text{for } a = 0 \end{cases}$$

where $\alpha \in (0, 2]$ denotes the impulsiveness, $\beta \in [-1, +1]$ the skewness, $\gamma > 0$ the scale parameter for dispersion and μ the location parameter, which can be seen as the equivalent to the mean value in a Gaussian distribution.

The statistical process of integrating the TAC in every voxel and the noise in the SPECT images lead to outliers which are contained in the VOIs. The Mahalanobis distance was used to identify and remove those outliers [32, 118]. This distance measure is scale-invariant, unit-less, and takes into account the two-point correlations of the data set. To include only dose values that belong to the assumed statistic, all values with a Mahalanobis distance smaller than one were selected. The resulting location parameter for the alpha stable distribution was used to represent each VOI.

All patients have more than one lesion, therefore a TTD was calculated as follows:

$$TDD = \frac{\sum_{i=1}^{n} \langle D_{tumor,i} \rangle \cdot V_{tumor,i}}{\sum_{i=1}^{n} V_{tumor,i}}$$

where V_{tumor} is the tumor volume, $\langle D_{tumor} \rangle$ the mean tumor dose value per lesion and n the number of lesions.

9.6.2 Network architecture

The network architecture for the segmentation neural network is chosen equal to the one described in chapter 9.4.2. The only difference are the output data and the up-sampling process. The input are the dose distribution calculated by the soft scaling method and the IMFs from the CT image. The output in this case has three dimensions. Where the third dimensions has 4 slices. The first one has the left kidney segmentation, the second the right kidney segmentation, the third one the spleen segmentation and the last one the tumor segmentation. All the segmentation results refer to the center slice of the five slices from the input images.

9.6.3 Results

The neural network was evaluated with a LOOCV. The segmentations for the right kidney, the left kidney, the spleen and the tumors were estimated by the neural network based on the dose estimation of a soft tissue kernel and the CT image. As ground truth a manual segmentation was used.

In figure 9.14, the manual and the estimated segmentation for one slice is shown. The manual segmentation is binary, but the estimated segmentation has values between 0 and 1. Hence, these values can be interpreted as occurrence probabilities. Considering the estimated segmentation, two segments for the kidney are clearly visible. One of the segments is much smaller than the other and

Table 9.8: Absolute mean values as well as min-max values per VOI for all 26 patients

region of interest (%)	mean	min	max
left kidney	02.80	0.01	12.69
right kidney	03.23	0.01	09.70
spleen	04.00	0.14	20.60
tumor	11.19	0.32	63.27

also has a smaller probability. The same can be seen for the spleen segmentation in this slice. These artifacts have to be removed. For this reason, after the estimation of the segmentation for the kidneys and the spleen, all connected regions were detected and only the largest one was chosen. With this method, all additional regions outside the organ of interest could be removed.

Figure 9.14: On the left side the manual segmenation is illustrated and on the right side the estimated one. In both images in the left top corner the left kidney is shown, in the right top corner the right kidney, in the left bottom corner the spleen and in the right bottom corner the tumor lesions.

For each segmentation, the mean doses were calculated based on full Monte Carlo simulations. For each VOI, the mean dose was also computed based on a manual segmentation. The latter and the estimated neural network segmentation (*Ground truth* (GT)) were compared to each other. The absolute and relative percentage differences between these two values for four regions of interest of 26 patients are illustrated in figure 9.15.

The absolute mean values, the minimum and maximum per VOI of all 26 patients are collected in the following table 9.9.

All in all, the results are quite good. They could be improved by a larger number of training data, though. The kidney and spleen segmentation were very

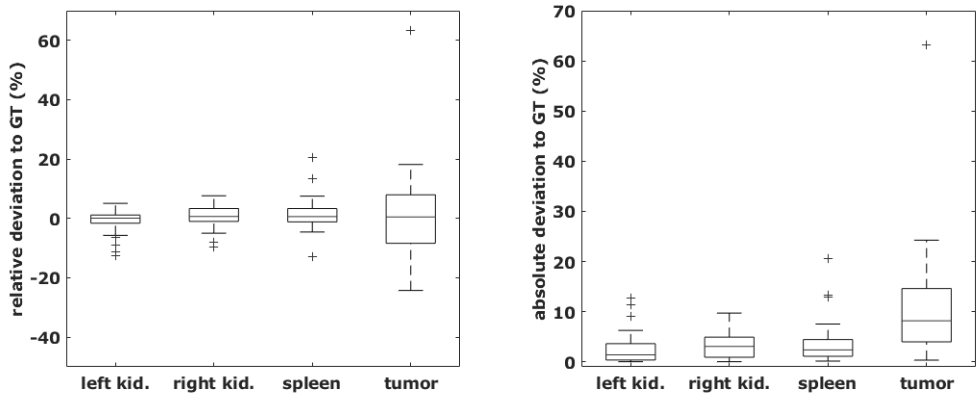


Figure 9.15: Relative (left side) and absolute (right side) deviation between mean dose value averaged over the manuall segmenation and the segmentation estimated by the neural network.

easy for a human based on the CT image, because the interobserver variability was small. But the tumor segmentation had a much larger variability, hence had to be performed much more individually. Therefore, the automated tumor segmentation was very hard.

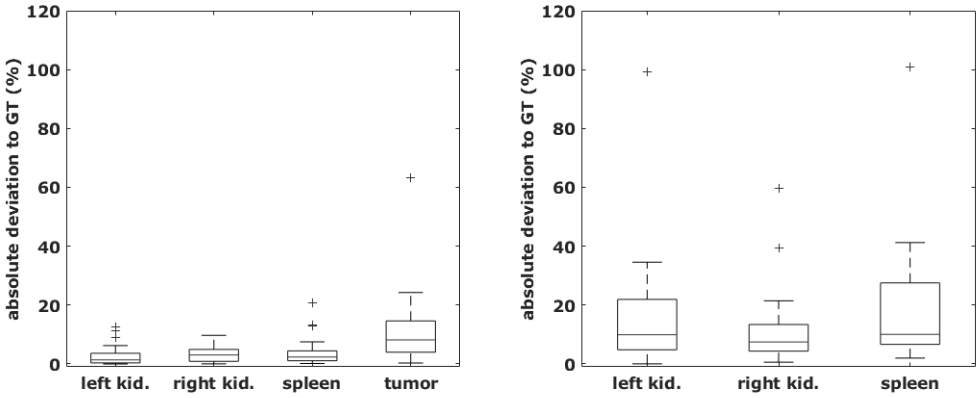


Figure 9.16: Absolute deviation between mean dose value averaged over the manuall segmenation and the segmentation estimated by the neural network. On the left side for the SPECT-CT-NN and on the left side the etablished NN.

The results of the neural network were compared with an established segmentation neural network. This network used the raw CT data and estimated the kidneys and the spleen. A tumor segmentation was not possible because of the fact that tumors were not visible in a low dose CT image. For each segmentation, the

Table 9.9: Mean percentage dose deviation as well as min-max values, averaged over all patients

region of interest	mean	min	max
left kidney	29.54	1.24	344.13
right kidney	11.21	0.21	53.55
spleen	18.18	1.44	85.42

mean dose per organ was calculated. The percentage deviations between the dose values of the neural network segmentation and the manual segmentation were determined. The percentage values are illustrated in figure 9.16 on the right side. There is also one outlier for the left kidney with a deviation of 344% which cannot be seen in the images. The mean, minimum and maximum values, averaged over all patients, are summarized in the table 9.9.

Chapter 10

Medical Achievements

In this chapter, the influence of the applied dose to the tumor lesion upon the therapy results will be evaluated. The ^{177}Lu therapy is established for patients suffering from either PC or from neuroendocrine tumors (NETs). For 13 patients that underwent ^{177}Lu DOTATOC therapy and 13 patients that underwent ^{177}Lu PSMA therapy, dose values were computed for every voxel separately. These dose values were then correlated with different clinical parameters.

10.1 Statistical analysis

The Spearman correlation coefficient was used to determine the significance of relations. To quantify differences between two groups, a Wilcoxon rank-sum test was applied. For all analyses, a p-value < 0.05 was considered statistically significant. The confidence intervals of binary variables were determined by assuming a binomial distribution.

10.2 PSMA

This section is abstracted from [TIG-7]. The clinical parameters from all patients are collected in table 6.2. The calculated dose values for the spleen, kidneys and the tumor are summarized in table 10.1.The mean half-life for the kidney was $38.2 \pm 16.6\ h$ while tracer-positive tumor lesions were characterized by a half-life of $34.9 \pm 135.1\ h$. The values are in agreement with other studies [33, 161, 71].

The Gleason Score is the grading system used to determine the aggressiveness of prostate cancer. Typical Gleason Scores range from $6 - 10$. The higher the Gleason Score, the more likely the cancer will grow and spread quickly. The TTD was compared between Gleason Scores ≤ 8 and > 8, and a trend towards higher TTD values with higher Gleason Scores has been observed (see figure 10.1 left

Table 10.1: Dose values for all ROIs.

organ	mean and standard deviation (Gy)	range (Gy)
kidney	2.59 ± 0.63	$1.67 - 3.92$
spleen	0.79 ± 0.46	$0.31 - 1.90$
tumor	11.00 ± 11.97	$1.28 - 49.10$

side). Also a trend towards a higher TTD for lymph node metastases compared to bone metastases was observed (see figure 10.1 right side). However, statistical significance was not reached, most probably due to the small number of patients in our study.

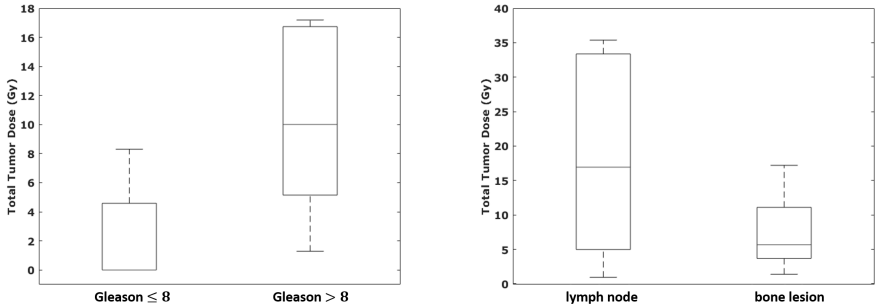


Figure 10.1: A trend towards higher TTDs was observed in patients with initial Gleason Scores of 8 or higher compared to patients with a Gleason Score below that threshold (left side). TTD of PSMA-positive lymph node metastases in comparison with PSMA-positive bone metastases (right side).

PSA is a protein produced by normal, as well as malignant, cells of the prostate gland. The PSA value measured in the blood can be used as good tumor marker for prostate cancer patients. Therefore, the correlation between the serum-PSA level before therapy and the following parameters was tested:

- TTD
- Tumor volume
- Total number of lesions.

A significant correlation between serum PSA levels before therapy and the tumor load, represented by the number of PSMA-positive lesions ($r = 0.56, p <$

0.05) and the volume of PSMA-positive lesions ($r = 0.69, p < 0.05$) was found. A higher serum-PSA level was associated with a higher tumor load and a higher number of lesions. The serum-PSA level has no significant correlation with the TTD. This is in line with previous results where a higher PSMA expression and a higher tumor load was seen with increasing tumor aggressiveness and impaired patient outcome [120, 4].

Furthermore, the absolute and percentage differences in PSA-serum level before and after therapy were analyzed. Both parameters were tested for significance with

- TTD
- Tumor volume
- Total number of lesions.

A significant correlation between the TTD with the percentage change of serum-PSA levels before and after therapy was observed (Spearman, $r = -0.57$, $p < 0.05$, see figure 10.2). Since the time between initiation and completion of the therapy differed in our patient collective, percentage changes of serum PSA values per month were calculated and yielded a significant correlation to the TTD (r= -0.59, p<0.05). Higher TTDs were associated with a decrease of serum PSA levels, while lower TTD were associated with an increase of serum PSA levels. This is an important finding since assessment of biochemical response as measured by PSA serum levels, is also recommended for assessment of many therapies in advanced prostate cancer by current guidelines [58] and assessment of TTD during multiple cycles might be a predictor of response to therapy.

No significant correlation could be found between the percentage change of PSA-serum levels and the number and volume of tumor lesions.

In patients with higher total tumor volumes of PSMA-positive lesions, kidney average doses were significantly lower compared to patients with small total tumor volumes ($r = -0.58, p < 0.05$) (see Figure 10.2). In radio-ligand therapy with somatostatin receptor type 2–specific peptides and for PSMA PET/CT, it has been shown that tumor burden significantly affects the activity concentration in tumor tissue and in dose-limiting organs [12, 46, 13, 77]. These results suggest that with an increasing tumor load the biologically effective doses to tumor and normal tissue, and also the tumor to kidney ratio decreased. These results might be explainable by a reduction of renal uptake with less available peptide. The ratio could be improved by applying larger amounts of peptide with correspondingly adapted activity. However, so far, in patients with metastatic castration-resistant prostate cancer that underwent ^{177}Lu-PSMA-*Radioligand Therapy* (RLT) standard activities, and amounts of peptide are usually injected despite the considerably interindividual differences in the amount of tumor burden. Large tumor

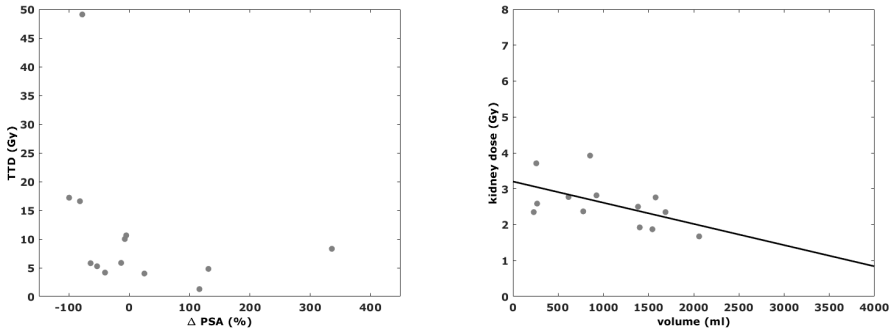


Figure 10.2: Higher Total Tumor Doses in PSMA-positive tumor lesions correlated significantly with the percentage change in PSA-serum level (seft side). Estimated kidney doses show an inverse correlation with the volume of PSMA-positive tumor lesions (right side).

burdens might therefore lead to a less effective treatment if standard activities are used. Further, larger ideally prospective multicenter trials with long-term follow-up are necessary to determine whether this novel Monte Carlo based voxel-wise dosimetry approach in combination with the histological, biochemical, and clinical parameters that were evaluated in this study might have an influence on individualization of further therapy planning and on the adaption of administered ^{177}Lu-PSMA activities.

10.3 DOTATOC

This section is abstracted from [TIG-8]. The information about the patients are collected in table 6.1.

In figure 10.3, average dose values for subgroups of tracer-positive tumor lesions and representative fused axial SPECT/CT images are illustrated. The average dose per cycle was

- for the spleen $0.67 \pm 0.41\ mGy/MBq (0.20 - 1.62\ mGy/MBq)$,
- for the kidneys $0.52 \pm 0.20\ mGy/MBq (0.30 - 0.97\ mGy/MBq)$ and
- for all 198 tracer-positive tumor lesions $1.46 \pm 1.26\ mGy/MBq (0.20 - 5.60\ mGy/MBq)$.

The dose values for the spleen and kidneys are in agreement with the value of $0.6\ mGy/MBq$ and of $0.7\ mGy/MBq$ respectively, averaged over 59 patients and published in [140]. Average injected activity was $6532 \pm 449\ MBq\ (5776 -$

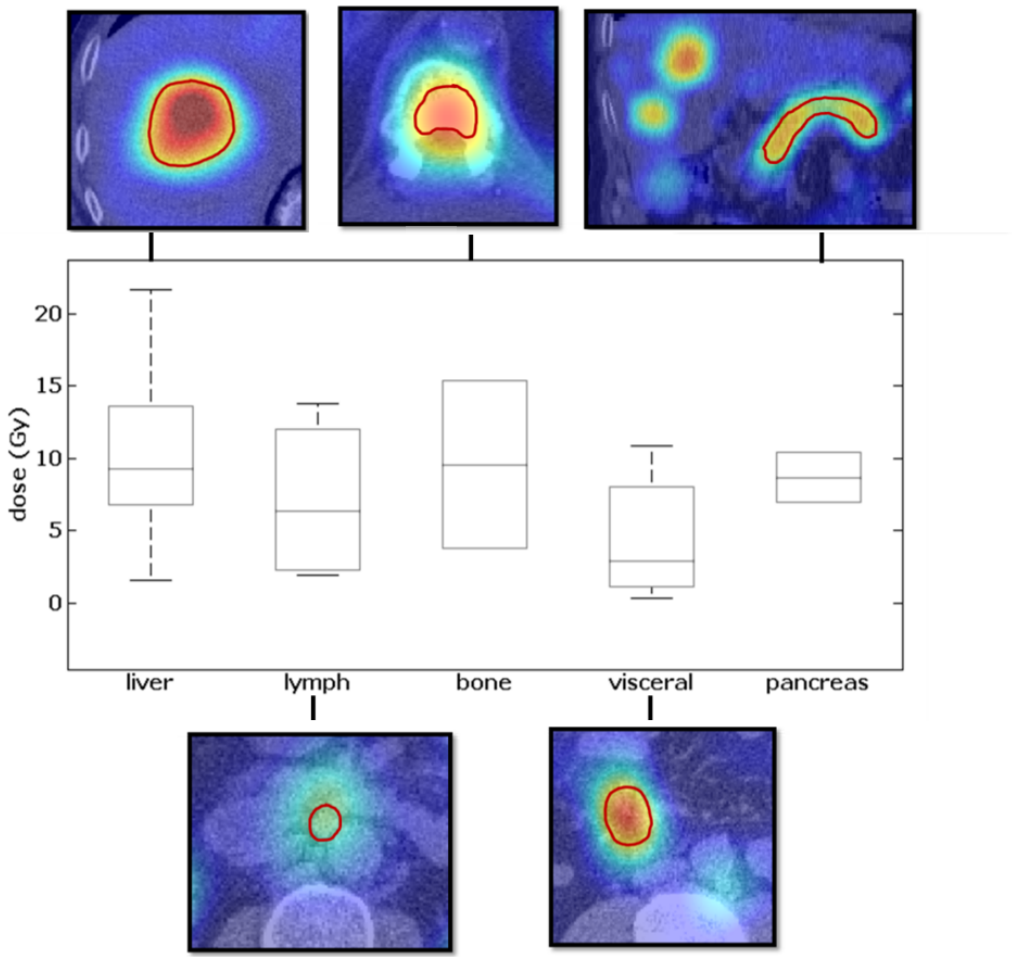


Figure 10.3: Averaged dose values for different kind of metastases.

$7264\ MBq$). The mean halflife for the tumor lesions was $61.3\ h \pm 117.8\ h$ and for the kidney $67.4\ h \pm 20.0\ h$. The mean effective half-lives of ^{177}Lu-DOTATOC in the kidneys are very similar to the values of $63\ h$, averaged over 30 patients and reported in [118].

Neuroendocrine tumors have an own tumor classification. Neuroendocrine tumors were graded as G1, G2, or G3 based on mitotic count and Ki-67 labeling index and presence of necrosis. Well differentiated tumors are graded into G1, while intermediate and high grade tumors are graded in G2 and G3 respectively. Low and intermediate grade tumors (G 1-2) absorbed a higher TTD compared to high grade tumors (G 3) (Signed-rank-test, p= <0.05) (see Figure 10.3). Most studies reported in the literature are conducted using ^{177}Lu-DOTATATE PRRT and are limited to well differentiated (G1-G2) tumors, and only a small portion

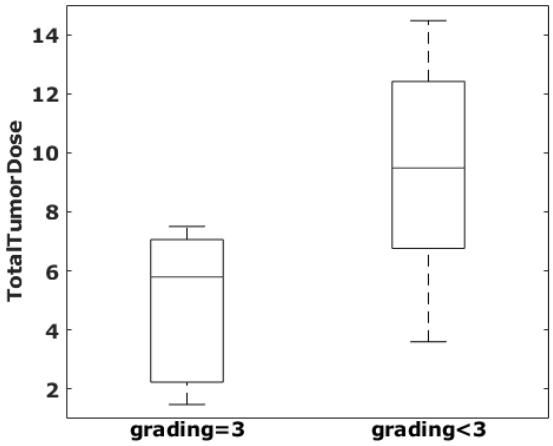


Figure 10.4: TTD for different tumor gradings.

included G3 tumors [24, 25], which are characterized by a short overall survival of $4 - 6$ months [155]. The reason for this is probably the higher expression of somatostatin receptors in well differentiated tumors.

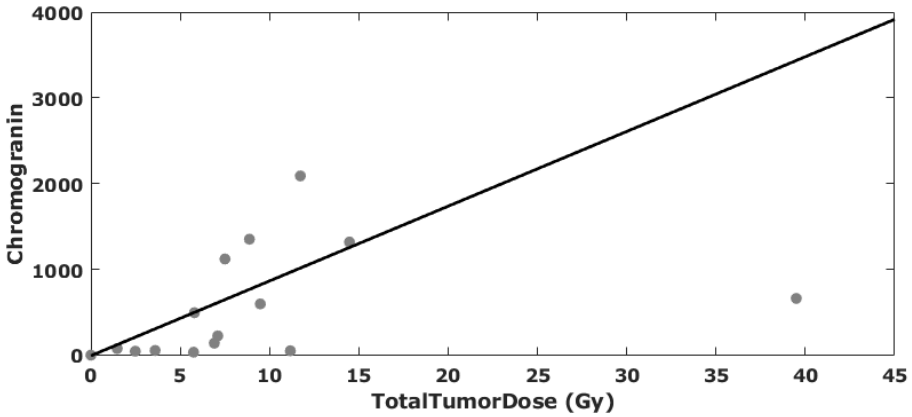


Figure 10.5: Chromogranin before therapy in dependence of the TTD.

The plasma chromogranin A (CgA) level is a reliable biomarker for identifying patients with advanced gastroenteropancreatic neuroendocrine tumors. The CgA value before therapy correlated significantly with the TTD (Pearson-corr.: $\rho = 0.67$, $p = 0.01$) (see figure 10.5). A higher CgA value resulted in a higher TTD.

TTD values also significantly correlated with the difference between CgA val-

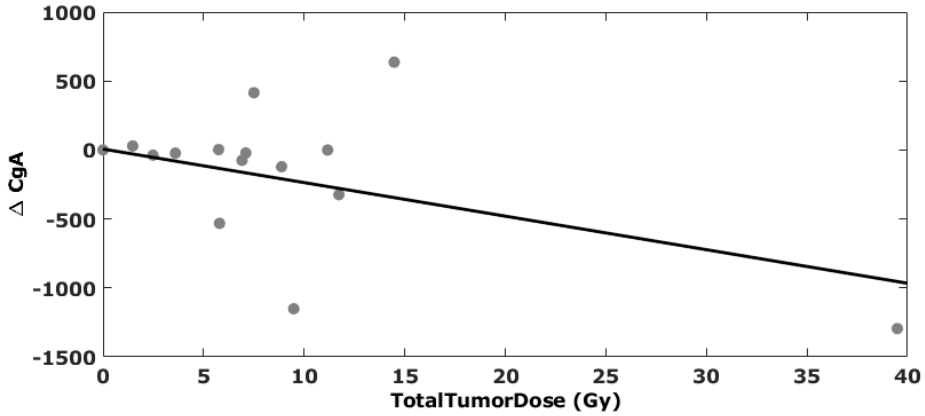


Figure 10.6: Difference of chromogranin in dependence of the TTD.

ues measured before and after therapy (Pearson-correlation: $\rho = -0.54, p = 0.0451$) (see Figure 10.6).
It has been proposed that CgA is more frequently elevated in well-differentiated tumors compared to poorly differentiated tumors of the midgut [104]. Furthermore, CgA is valuable in evaluating the efficacy of a broad range of therapies in NETs, including sandostatin therapy [151] or PRRT [103]. We could demonstrate that higher pre-therapeutic CgA values significantly correlated with the TTD, and a higher TTD resulted in a significant decrease of CgA values. However, clinicians should be aware that an increase of CgA values following ^{177}Lu-PRRT-therapy might be observed even in patients with an objective response or stable disease [156].

10.4 Limitations

Our studies suffer from several limitations. First of all, results should be interpreted with caution due to the small number of patients. Furthermore, this analyses were conducted as single center studies in a retrospective fashion. Also, the retrospective nature of these analyses have typical limitations, including possible biases stemming from patient referrals and treatments. Long-term follow-up to determine response to therapy would have been preferable, but was not feasible. Due to the manual fashion of the VOI definition exact position and size of the VOI are subject to intra- and inter-observer variability. Since the routinely acquired SPECT/CT dosimetry images at our institution cover the abdomen and pelvis, radiation dose to the parotid could not be calculated.

Conclusion

In this study, different dosimetry methods of ^{177}Lu therapies for neuroendocrine tumors and prostate cancer were evaluated. During such radionuclide therapies, the patient gets injected a radioactive substance, which is taken up tumor cells. Four hours after injection, the first SPECT-image will be acquired. Subsequent images are taken at 24h, 48h and 72h after injection. From the image intensities, the activity of each voxel can be deduced and similarly a density information can be obtained from the CT images. Medical experts are interested in the dose, which is absorbed by the kidneys, because this is the limiting organ.

According to the standard MIRD protocol, the dose applied to the kidneys is estimated as the product of two factors: a TIA and a dose kernel, called S-factor. The TIA in the kidney is computed by approximating a time series of measured SPECT intensities by a mono-exponential function, which is integrated over time. The resulting number of nuclear disintegrations (number of decays) are multiplied by an S-value, which can be determined through a Monte Carlo simulation of a standard phantom. After scaling the resulting dose value by the patient's weight and kidney volume, a dose value for the kidney can be obtained. However, as this standard MIRD method strongly depends on the medical expert, the error of this method amounts to roughly 25.0%. The aim of this study was to improve dose estimation via an objective computation without any dependence on a medical expert. Especially a voxel-wise dosimetry should be performed instead of an organ-wise dosimetry, and a patient-specific Monte Carlo simulation should replace the standard human phantom.

The first challenge of this work was to calculate the number of decays per voxel from the SPECT-images. Therefore, a statistical method called particle filter was used. With this method the voxel-wise noise of the SPECT images could be removed by matching the measurements to a physical model of the radioactive decay. This method yields a homogeneous distribution of the decays per voxel by taking into account different organs. The particle filter method was compared to a simple fit method. The latter generally resulted in inhomogeneous decay maps where certain voxels contained diverging numbers of decays. For 26 patients suffering from a prostate cancer or a neuroendocrine tumor, the corresponding

decay maps could be determined reliably.

In the main chapter of the work, different dose estimation methods were compared. As ground truth, a full Monte Carlo simulation of the absorbed dose per decay was used based on CT-derived density maps of all patients. Such Monte Carlo simulations are very time consuming, however, and thus not integrable into daily clinical routine. For this reason, fast methods for an accurate estimation of the dose distribution are of special interest to physicians. One such approximate method for dose estimation is a convolution of an activity map with simulated dose-voxel-kernels (DVK). To compute the latter, a homogeneous density kernel is implemented in a Monte Carlo simulation with a radioactive isotope in its center. Then a large number of nuclear decays needs to be simulated to obtain an accurate statistical dose distribution within this density kernel. If finally an activity map is convolved with this DVK, a dose distribution will be the result.

In literature there are a few methods published to convolve activity maps with rescaled DVKs. The re-scaling is done in various ways according to the mass density of the voxel, because absorbed dose strongly depends on density. In this respect, each local convolution of the activity map with a corresponding DVK would need a different kernel. But one DVK has the size of $9 \times 9 \times 9$ voxels, which encompasses 729 voxels altogether. As the mass density of the voxels varied between 0.5 to $2.0\ \frac{g}{cm^3}$ with a step size of $0.01\ \frac{g}{cm^3}$, this would have resulted in almost $150^{729} = 2.3 \cdot 10^{1586}$ different kernels, even neglecting rotational degrees of freedom. The MC of this huge number of kernels is too time-consuming to be performed. Therefore, a neural network was designed to predict the DVK from a given density kernel. The results are much better than the established scaling methods.

With the neural network method for estimation of a DVK, only the kernels from the kidneys could be predicted because of computational time. To obtain a dose distribution of a whole patient a second neural network was built to predict the dose of whole patients. The network yielded results which are in a similar range to the statistical error of the MC.

In the last sub-chapter two automatic segmentation methods were evaluated. The kidneys, spleen and tumors were segmented automatically by an established neural network using only the CT image and a second method which used the CT- and SPECT-image. The results of the latter are much better than for the first one, because tumors can only be detected in the SPECT-image.

In the last part of this study the medical achievements were explained. Through the appropriate dosimetry method a dose value could be determined which is in good correlation with clinical parameters. For both therapies, predicting parameters for the dose could be identified. The method enables a patient specific tumor therapy.

Publications based on this study

Based on this study the following paper were written:

TIG-1 **Factors affecting accuracy of S values and determination of time-integrated activity in clinical ^{177}Lu dosimetry**
Götz TI, Schmidkonz C, Lang EW, Maier A, Kuwert T, Ritt P
Annals of Nuclear Medicine, 2019 May 22

TIG-2 **Particle Filter De-noising of Voxel-specifc Time-Activity-Curves in Personalized ^{177}Lu Therapy**
Götz TI, Lang EW, Schmidkonz C, Maier A, Kuwert T, Ritt P
Zeitschrift für medizinische Physik, 2019 Dec 16

TIG-3 **A Comparison of Methods for Adapting ^{177}Lu Dose-Voxel-Kernels to Tissue Inhomogeneities**
Götz TI, Schmidkonz C, Lang EW, Maier A, Kuwert T, Ritt P
Physics in Medicine and Biology, 2019 Dec 19

TIG-4 **Neural Networks for Estimation of Dose Voxel Kernels**
Götz TI, Lang EW, Schmidkonz C, Kuwert T, Ludwig B
Zeitschrift für medizinische Physik, submitted

TIG-5 **A Deep Learning Approach to Radiation Dose Estimation**
Götz TI, Schmidkonz C, Chen S, Al-Baddai S, Kuwert T, Lang EW
Physics in Medicine and Biology, 2020 Feb 4

TIG-6 **Organwise Radiation Dose Estimation - A Deep Learning Approach**
Götz TI, Schmidkonz C, Chen S, Kuwert T, Lang EW, Ludwig B
European Journal of Nuclear Medicine and Molecular Imaging - Physics, submitted

TIG-7 **Three-dimensional Monte Carlo-based voxel-wise tumor dosimetry in patients with neuroendocrine tumors who underwent ^{177}Lu-DOTATOC therapy**
Götz TI, Lang EW, Prante O, Maier A, Cordes M, Kuwert T, Ritt P, Schmidkonz C
Annals of Nuclear Medicine, 2020 Feb 29

TIG-8 **Estimation of ^{177}Lu-PSMA-617 tumor uptake based on three-dimensional Monte Carlo based voxel-wise tumor dosimetry in patients with metastasized castration resistant prostate cancer**
Götz TI, Lang EW, Prante O, Cordes M, Kuwert T, Ludwig B, Ritt P, Schmidkonz C
Annals of Nuclear Medicine, 2020 Sep 1

List of Figures

1.1 Steps of the dosimetry chain according to the MIRD protocol . . . 20
1.2 Steps of a voxel-wise dosimetry chain 21

4.1 The figure shows the three splines along with their tension values that pass through the same extremal (local maxima and minima) points. Note that when the tension vanishes $T = 0$, the points are connected by a cubic spline; and when $T = 1$, it is representing a linear spline. 67

5.1 Convolutive filtering 73
5.2 Zero padding 74
5.3 Sobel edge detector 76
5.4 On the left side the gradient of a logistic activation function is illustrated and on the right side of the ReLU. 77
5.5 An illustration of a typical U-Net architecture. 81

6.1 CT-image, SPECT-image and hyprid SPECT/CT-image as shown in ITK-snap. In the right bottom corner the segmentation of the right and left kidney, the spleen and a liver tumor can be seen. . . 90
6.2 The four SPECT-images, the CT-image and the corresponding segmentation map illustrated for one slice of one patient. 91

8.1 Time activity curve for one kidney: $f(0) = 1\%IA, T_{1/2} = 77.9h$. . 100
8.2 Left side two histograms with a Kullback-Leibler divergence of 0.67 and on the right side with 0.12. 101
8.3 Variation of the number of particles. 103
8.4 Time activity curve vor two different voxels at four time steps. . . 104
8.5 One slice of the results for the particle filtering method (left) and the simple fit method (right). 105
8.6 Histograms of number of decays per voxel for the particle filter method (blue) and the voxel-wise simple fit of the four activity values (red). 105

8.7 Histograms of halflives per voxel for the particle filter method (blue) and the voxel-wise simple fit of the four activity values (red). The halflife for the whole organ is $81.6h$. 106

9.1 Relative error in percent resulting for one voxel from DVK convolution computations. 112
9.2 On the left side, the result of a MC for one patient is shown. The colorbar reflects an encoding of the deposited dose in $[Gy/decay]$. On the right side, the rSE for the MC is shown. The colors give information about the error in $\%$. 113
9.3 Percentage deviation of total dose values of the left kidney (left side) and right kidney(right side) between three methods for 24 patients. 114
9.4 Network architecture, where each convolutional layer includes a leaky ReLu activation layer and a batch normalization layer. . . . 118
9.5 On the left side, a CT image from the left kidney, divided in density kernels, one of which is illustrated in the middle. With a MC, the DVK has been calculated (seen on the right side). The MC should be replaced by a NN 120
9.6 Training and validation loss function in dependence of the number of epochs. 122
9.7 Percentage deviation between MC and the two methods in dependence on the density of the center voxel of the density kernel. The occurrence of the density within all kidneys can also be seen on the ordinate. 122
9.8 Selection of 5 slices from one patient used for estimating the dose of the center slice. 123
9.9 Net architecture with two 3D inputs and one 2D output. 124
9.10 The six IMFs for a patient suffering from a prostate cancer. 125
9.11 Differences in dose for right kidneys compared between seven methods. 130
9.12 Percentage deviations ($\%$) for four different methods ordered according to the tissue class lung tissue, and cortical bone. 131
9.13 Histogram of all dose values for a PSMA-positive bone metastasis. 132
9.14 On the left side the manual segmenation is illustrated and on the right side the estimated one. In both images in the left top corner the left kidney is shown, in the right top corner the right kidney, in the left bottom corner the spleen and in the right bottom corner the tumor lesions. 134

9.15 Relative (left side) and absolute (right side) deviation between mean dose value averaged over the manuall segmenation and the segmentation estimated by the neural network. 135
9.16 Absolute deviation between mean dose value averaged over the manuall segmenation and the segmentation estimated by the neural network. On the left side for the SPECT-CT-NN and on the left side the etablished NN. 135

10.1 A trend towards higher TTDs was observed in patients with initial Gleason Scores of 8 or higher compared to patients with a Gleason Score below that threshold (left side). TTD of PSMA-positive lymph node metastases in comparison with PSMA-positive bone metastases (right side). 138
10.2 Higher Total Tumor Doses in PSMA-positive tumor lesions correlated significantly with the percentage change in PSA-serum level (seft side). Estimated kidney doses show an inverse correlation with the volume of PSMA-positive tumor lesions (right side). . . . 140
10.3 Averaged dose values for different kind of metastases. 141
10.4 TTD for different tumor gradings. 142
10.5 Chromogranin before therapy in dependence of the TTD. 142
10.6 Difference of chromogranin in dependence of the TTD. 143

Bibliography

[1] A. F. Agarap. Deep learning using rectified linear units (relu), 2019.

[2] Ch. C. Agarwal. *Neural Networks and Deep Learning*. Springer, 2018.

[3] H. Ahmadzadehfar, E. Eppard, S. Kürpig, R. Fimmers, A. Yordanova, C. Schlenkhoff, F. Gärtner, S. Rogenhofer, and M. Essler. Therapeutic response and side effects of repeated radioligand therapy with 177Lu-PSMA-DKFZ-617 of castrate-resistant metastatic prostate cancer. *Oncotarget*, 7(11):12477, 2016.

[4] H. Ahmadzadehfar, S. Schlolaut, R. Fimmers, A. Yordanova, S. Hirzebruch, C. Schlenkhoff, and et al. Predictors of overall survival in metastatic castration-resistant prostate cancer patients receiving [177Lu] Lu-PSMA-617 radioligand therapy. *Oncotarget*, 8:103108, 2017.

[5] S. Al-Baddai, K. Al-Subari, A.M. Tomé, J. Solé-Casals, and E.W. Lang. A Green's function-based Bi-dimensional Empirical Mode Decomposition. *Information Sciences*, 348:305–321, 2016.

[6] S. Al-Baddai, K. Al-Subari, A.M. Tomé, G. Volberg, S. Hanslmayr, R. Hammwöhner, and E. W. Lang. Bidimensional ensemble empirical mode decomposition of functional biomedical images taken during a contour integration task. *Biomedical Signal Processing and Control*, 13:218–236, 2014.

[7] M.S. Anbarasi, S. Ghaayathari, R. Kamaleswari, and I. Abirami. Outlier Detection for Multidimensional Medical Data. *International Journal of Computer Science and Information Technologies*, 2:512–516, 2011.

[8] J.K. Annkah, I. Rosenberg, N. Hindocha, S.A. Moinuddin, K. Ricketts, A. Adeyemi, and G. Royle. Assessment of the dosimetric accuracies of CATPhan 504 and CIRS 062 using kV-CBCT for performing direct calculations. *Journal of medical physics/Association of Medical Physicists of India*, 39(3):133, 2014.

[9] V. Arati, M. D. Rao, G. Akabani, and D. A. Rizzieri. Radioimmunotherapy for Non-Hodgkin's Lymphoma. *Clinical Medicine and Research*, 3(3):157–165, 2005.

[10] P. Arce, J.I. Lagares, L. Harkness, D. Pérez-Astudillo, M. Cañadas, P. Rato, M. de Prado, Y. Abreu, G. de Lorenzo, M. Kolstein, et al. Gamos: A framework to do Geant4 simulations in different physics fields with an user-friendly interface. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 735:304–313, 2014.

[11] R. Barone, F. Borson-Chazot, and R. et al. Valkema. Patient-specific dosimetry in predicting renal toxicity with ^{90}Y-DOTATOC: Relevance of kidney volume and dose rate in finding a dose-effect relationship. *Journal of Nuclear Medicine*, 46:99S–106S, 2005.

[12] J.M. Beauregard, M.S. Hofman, G. Kong, and R.J. Hicks. The tumour sink effect on the biodistribution of 68Ga-DOTA-octreotate: implications for peptide receptor radionuclide therapy. *European Journal of Nuclear Medicine and Molecular Imaging*, 39(1):50–56, Jan 2012.

[13] N. J. Begum, A. Thieme, N. Eberhardt, R. Tauber, C. D'Alessandria, A.J. Beer, G. Glatting, M. Eiber, and P. Kletting. The effect of total tumor volume on the biologically effective dose to tumor and kidneys for 177Lu-labeled PSMA peptides. *Journal of Nuclear Medicine*, 59(6):929–933, 2018.

[14] M.J. Berger. Monte Carlo Calculation of the penetration and diffusion of fast charged particles. In *Methods in Comput. Phys., B. Alder and S. Fernbach and M. Rotenberg, eds.*, pages 135–215, New York, 1963. Academic Press.

[15] S.M.A. Bhuiyan, R.R. Adhami, and J.F. Khan. A novel approach of fast and adaptive bidimensional empirical mode decomposition. In *In IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, pages 1313 –1316, 2008.

[16] S.M.A. Bhuiyan, J.F. Khan, N.O. Attoh-Okine, and Adhami R.R. Study of Bidimensional Empirical Mode Decomposition Method for Various Radial Basis Function Surface Interpolators. In *2009 International Conference on Machine Learning and Applications*, pages 18–24. IEEE, 2009.

[17] A.F. Bielajew and D.W.O. Rogers. Electron Step-Size Artefacts and PRESTA. In *Monte Carlo Transport of Electrons and Photons, T. M. Jenkins and W. R. Nelson and A. Rindi ad A. E. Nahum and D. W. O. Rogers, eds.*, pages 115–137, New York, 1988. Plenum Press.

[18] V. Bochkarev, G. Radzievsky, L. Timofeev, and N. Demianov. Distribution of absorbed energy from a point beta-source in a tissue-equivalent medium. *The International journal of applied radiation and isotopes*, 23(11):493–504, 1972.

[19] L. Bodei, M. Cremonesi, and M. et al. Ferrari. Long-term evaluation of renal toxicity after peptide receptor radionuclide therapy with ^{90}Y-DOTATOC and ^{177}Lu-DOTATATE: The role of associated risk factors. *European Journal of Nuclear Medicine and Molecular Imaging*, 35:1847–1856, 2008.

[20] L. Bodei, M. Cremonesi, C. M. Grana, N. Fazio, S. Iodice, S. M. Baio, M. Bartolomei, D. Lombardo, M. E. Ferrari, M. Sansovini, et al. Peptide receptor radionuclide therapy with 177 Lu-DOTATATE: the IEO phase I-II study. *European journal of nuclear medicine and molecular imaging*, 38(12):2125–2135, 2011.

[21] W.E. Bolch, Bouchet L.G., and Robertson J.S. et al. MIRD pamphlet no. 17: the dosimetry of nonuniform activity distributions — Radionuclide S-values at the voxel level. *Journal of Nuclear Medicine*, 40:11S–36S, 1999.

[22] F. Botta, A. Mairani, G. Battistoni, M. Cremonesi, A. Di Dia, A. Fasso, A. Ferrari, M. Ferrari, G. Paganelli, G. Pedroli, et al. Calculation of electron and isotopes dose point kernels with FLUKA Monte Carlo code for dosimetry in nuclear medicine therapy. *Medical physics*, 38(7):3944–3954, 2011.

[23] A. Bräuer, L.S. Grubert, W. Roll, A.J. Schrader, M. Schäfers, M. Bögemann, and K. Rahbar. 177Lu-PSMA-617 radioligand therapy and outcome in patients with metastasized castration-resistant prostate cancer. *European journal of nuclear medicine and molecular imaging*, 44(10):1663–1670, 2017.

[24] V.D. Calhoun, T. Adali, L.K. Hansen, J. Larsen, and J.J. Pekar. ICA of Functional MRI Data: An Overview. In *in Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, pages 281–288, 2003.

[25] M.E. Caplin, M. Pavel, J.B. Ćwikła, A.T. Phan, M. Raderer, E. Sedláčková, G. Cadiot, E.M. Wolin, J. Capdevila, L. Wall, et al. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. *New England Journal of Medicine*, 371(3):224–233, 2014.

[26] S. Chiavasse, I. Aubineau-Laniece, A. Bitar, A. Lisbona, J. Barbet, D. Franck, J.R. Jourdain, and M. Bardies. Validation of a personalized dosimetric evaluation tool (Oedipe) for targeted radiotherapy based on the Monte Carlo MCNPX code. *Physics in Medicine and Biology*, 51:601–616, 2006.

[27] S. Chiavasse, M. Bardi, F. Guiraud-Vitaux, D. Bruel, J. R. Jourdain, D. Franck, and I. Aubineau-Lani. OEDIPE: a personalized dosimetric tool associating voxel-based models with MCNPX. *Cancer Biotherapy and Radiopharmaceuticals*, 20(3):325–332, 2005.

[28] A. Cichocki, S. Amari, K. Siwek, T. Tanaka, and A.H. et al. Phan. ICALAB Toolbox, 2007.

[29] A. Cichocki, R. Zdunek, A.H. Pham, and S. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley and Sons, 2009.

[30] M. Cremonesi, M. Ferrari, L. Bodei, G. Tosi, and G. Paganelli. Dosimetry in peptide radionuclide receptor therapy: a review. *Journal of nuclear medicine*, 47(9):1467–1475, 2006.

[31] Ch Damerval, S. Meignen, and V. Perrier. A Fast Algorithm for Bidimensional EMD. *IEEE Signal Processing Letters*, 12(10):701–704, 2005.

[32] R. De Maesschalck, D. Jouan-Rimbaud, and D.ésiré L Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1):1–18, 2000.

[33] A. Delker, W.P. Fendler, C. Kratochwil, A. Brunegraf, A. Gosewisch, F.J. Gildehaus, S. Tritschler, C.G. Stief, K. Kopka, U. Haberkorn, et al. Dosimetry for 177 Lu-DKFZ-PSMA-617: a new radiopharmaceutical for the treatment of metastatic prostate cancer. *European journal of nuclear medicine and molecular imaging*, 43(1):42–51, 2016.

[34] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum-Likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

[35] Y.K. Dewaraja, M.J. Schipper, P.L. Roberson, S.J. Wilderman, H. Amro, D.D. Regan, K.F. Koral, M.S. Kaminski, and A.M. Avram. ^{131}I-tositumomab radioimmunotherapy: initial tumor dose-response results using 3-dimensional dosimetry including radiobiologic modeling. *Journal of Nuclear Medicine*, 51:1155–1162, 2010.

[36] A. Dieudonné, R. Hobbs, W. Bolch, G. Sgouros, and I. Gardin. Fine-resolution voxel S values for constructing absorbed dose distributions at variable voxel size. *Journal of Nuclear Medicine*, 51(10):1600–1607, 2010.

[37] A. Dieudonné, R. Hobbs, R. Lebtahi, F. Maurel, S. Baechler, R. Wahl, A. Boubaker, D. Le Guludec, G. Sgouros, and I. Gardin. Study of the impact of tissue density heterogeneities on 3-dimensional abdominal dosimetry: comparison between dose kernel convolution and direct Monte Carlo methods. *Journal of Nuclear Medicine*, 54(2):236–243, 2013.

[38] R. Douc and O. Cappe. Comparison of resampling schemes for particle filtering. pages 64 – 69, 2005.

[39] A. Doucet, N. de Freitas, and N. (Eds.) Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[40] R.E. Drzymala, R. Mohan, L. Brewster, J. Chu, M. Goitein, W. Harms, and M. Urie. Dose-volume histograms. *International Journal of Radiation Oncology, Biology and Physics*, 21(1):71–78, 1991.

[41] A. Eberle. Markoc Processes, 2017.

[42] M. Fernández, H. Hänscheid, T. Mauxion, M. Bardiès, P. Kletting, G. Glatting, and M. Lassmann. A fast method for rescaling voxel s values for arbitrary voxel sizes in targeted radionuclide therapy from a single Monte Carlo calculation. *Medical physics*, 40(8), 2013.

[43] G.S. Fishman. *Monte Carlo - Concepts, Algorithms and Applications*. Springer, 1996.

[44] G. Flux, M. Bardies, and C. et al. Chiesa. Clinical radionuclide therapy dosimetry: The quest for the Holy Gray. *European Journal of Nuclear Medicine and Molecular Imaging*, 34:1699–1700, 2007.

[45] G. D. Flux, M. J. Guy, R. Beddows, M. Pryor, and M. A. Flower. Estimation and implications of random errors in whole-body dosimetry for targeted radionuclide therapy. *Physics in Medicine and Biology*, 47(7):3211–3223, 2002.

[46] F.C. Gaertner, K. Halabi, H. Ahmadzadehfar, S. Kürpig, E. Eppard, C. Kotsikopoulos, N. Liakos, R.A. Bundschuh, H. Strunk, and M. Essler. Uptake of PSMA-ligands in normal tissues is dependent on tumor load in patients with prostate cancer. *Oncotarget*, 8(33):55094, 2017.

[47] M. Garkavij, M. Nickel, K. Sjögreen-Gleisner, M. Ljungberg, T. Ohlsson, K. Wingårdh, S.E. Strand, and J. Tennvall. 177Lu-[DOTA0, Tyr3] octreotate therapy in patients with disseminated neuroendocrine tumors: Analysis of dosimetry with impact on future therapeutic strategy. *Cancer*, 116(S4):1084–1092, 2010.

[48] H.B. Giap, D.J. Macey, and D.A. Podoloff. Development of a SPECT-based three-dimensional treatment planning system for radioimmunotherapy. *Journal of Nuclear Medicine*, 36:1885–1894, 1995.

[49] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Massachusetts Institute of Technology Press, 2016.

[50] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-GaussianBayesian state estimation. *Radar and Signal Processing, IEEE Proceedings*, 140(2):107–113, 1993.

[51] Th. I. Götz, M. Ermer, D. Salas-González, M. Kellermeier, V. Strnad, Ch. Bert, B. Hensel, A.M. Tomé, and E.W. Lang. On the use of multi-dimensional scaling and electromagnetic tracking in high dose rate brachytherapy. *Physics in Medicine and Biology*, 62(20):7959, 2017.

[52] E. Grassi, F. Fioroni, V. Ferri, E. Mezzenga, M.A. Sarti, T. Paulus, N. Lanconelli, A. Filice, A. Versari, and M. Iori. Quantitative comparison between the commercial software STRATOS® by Philips and a homemade software for voxel-dosimetry in radiopeptide therapy. *Physica Medica: European Journal of Medical Physics*, 31(1):72–79, 2015.

[53] J. Grimes. *Patient-specific internal dose calculation techniques for clinical use in targeted radionuclide therapy*. PhD thesis, University of British Columbia, 2013.

[54] C. Gulcehre and Y. Bengio. Knowledge matters: importance of prior information for optimization. *J. Machine Learning Research*, 17:226–257, 2016.

[55] J. Gustafsson, G. Brolin, M. Cox, M. Ljungberg, L. Johansson, and K. Gleisner. Uncertainty propagation for SPECT/CT-based renal dosimetry

in 177Lu peptide receptor radionuclide therapy. *Physics in medicine and biology*, 60(21):8329, 2015.

[56] W.P. Harris, E.A. Mostaghel, P.S. Nelson, and B. Montgomery. Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. *Nature Reviews Urology*, 6(2):76, 2009.

[57] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[58] A. Heidenreich, P. Bastian, J. Bellmunt, Michel Bolla, S. Joniau, T. van der Kwast, M. Mason, V. Matveev, T. Wiegel, F. Zattoni, and N. Mottet. EAU Guidelines on Prostate Cancer. Part II: Treatment of Advanced, Relapsing, and Castration-Resistant Prostate Cancer. *European Urology*, 65(2):467 – 479, 2014.

[59] G. Hinton. RMSProp, 2016.

[60] R.F. Hobbs, T. McNutt, S. Baechler, B. He, C.E. Esaias, E.C. Frey, D.M. Loeb, R.L. Wahl, O. Shokek, and G. Sgouros. A treatment planning methodology for sequentially combining radiopharmaceutical therapy (RPT) and external radiation therapy (XRT). *International Journal of Radiation Oncology – Biology – Physics*, 80:1256–1262, 2011.

[61] R.F. Hobbs, R.L. Wahl, M.A. Lodge, M.S. Javadi, S.Y. Cho, D.T. Chien, M.E. Ewertz, C.E. Esaias, P.W. Ladenson, and G. Sgouros. ^{124}i PET-based 3D-RD dosimetry for a pediatric thyroid cancer patient: real-time treatment planning and methodologic comparison. *Journal of Nuclear Medicine*, 50:1844–1847, 2009.

[62] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, and H.H. Liu. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society London A*, 454:903–995, 1998.

[63] H.M. Hudson and R.S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Nuclear Science*, 13:601–609, 1994.

[64] C. Huggins and C.V. Hodges. Studies on prostatic cancer: I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. *The Journal of urology*, 168(1):9–12, 2002.

[65] A. Humeau-Heurtier, G. Mahe, and P. Abraham. Multi-dimensional complete ensemble empirical mode decomposition with adaptive noise applied to laser speckle contrast images. *Medical Imaging, IEEE Transactions on*, PP(99):1–1, 2015.

[66] A. Hyvärinen, J. Karhunen, and E. Oja. Independent Component Analysis. *John Wiley,New York*, 2001.

[67] E. Ilan, M. Sandstrom, C. Wassberg, A. Sundin, U. Garske-Román, B. Eriksson, et al. Dose response of pancreatic neuroendocrine tumors treated with peptide receptor radionuclide therapy using 177Lu-DOTATATE. *Journal of Nuclear Medicine*, 56(2):177–82, 2015.

[68] M. Isard and A. Blake. CONDENSATION– conditional density propagation of visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 2001.

[69] P. Jackson, J.M. Beauregard, M.S. Hofman, T. Kron, A. Hogg, and R.J. Hicks. An automated voxelized dosimetry tool for radionuclide therapy based on serial quantitative SPECT/CT imaging. *Medical physics*, 40(11), 2013.

[70] N.D. James, M.R. Sydes, N.W. Clarke, M.D. Mason, D.P. Dearnaley, M.R. Spears, A.W.S. Ritchie, C.C. Parker, J. M. Russell, G. Attard, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (stampede): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *The Lancet*, 387(10024):1163–1177, 2016.

[71] L. Kabasakal, M. AbuQbeitah, A. Aygün, N. Yeyin, M. Ocak, E. Demirci, and T. Toklu. Pre-therapeutic dosimetry of normal organs and tissues of 177 Lu-PSMA-617 prostate-specific membrane antigen (PSMA) inhibitor in patients with castration-resistant prostate cancer. *European journal of nuclear medicine and molecular imaging*, 42(13):1976–1983, 2015.

[72] R.E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal Basic Engineering*, pages 35–45, 1960.

[73] K. Kanazawa, D. Koller, and S. Russell. Stochastic Simulation Algorithms for Dynamic Probabilistic Networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, pages 346–351, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.

[74] T. Karantanos, P.G. Corn, and T.C. Thompson. Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. *Oncogene*, 32(49):5501, 2013.

[75] I. Kawrakow, E. Mainegra-Hing, D.W.O. Rogers, F. Tessier, and B.R.B. Walters. The EGSnrs Code System, 2013.

[76] M. Khazaee Moghadam, A. Kamali Asl, P. Geramifar, and H. Zaidi. Evaluating the Application of Tissue-Specific Dose Kernels Instead of Water Dose Kernels in Internal Dosimetry: A Monte Carlo Study. *Cancer Biotherapy and Radiopharmaceuticals*, 31(10):367–379, 2016.

[77] P. Kletting, A. Thieme, N. Eberhardt, A. Rinscheid, C. D'Alessandria, J. Allmann, H.J. Wester, R Tauber, A.J. Beer, G. Glatting, et al. Modeling and predicting tumor response in radioligand therapy. *Journal of Nuclear Medicine*, 60(1):65–70, 2019.

[78] D.S. Klimstra, I.R. Modlin, D. Coppola, R.V. Lloyd, and S. Suster. The pathologic classification of neuroendocrine tumors: a review of nomenclature, grading, and staging systems. *Pancreas*, 39(6):707–712, 2010.

[79] P. Knoll, D. Kotalova, G. Kochle, I. Kuzelka, G. Minear, S. Mirzaei, M. Samal, L. Zadrazil, and H. Bergmann. Comparison of advanced iterative reconstruction methods for SPECT/CT. *Zeitschrift für Medizinische Physik*, 22:58–69, 2012.

[80] K. S. Kolbert, G. Sgouros, A. M. Scott, J. E. Bronstein, R. A. Malane, J. Zhang, H. Kalaigian, S. McNamara, L. Schwartz, and S. M. Larson. Implementation and evaluation of patient-specific three-dimensional internal dosimetry. *Journal of Nuclear Medicine*, 38(2):301–308, 1997.

[81] S. Konishi and G. Kitagawa. *Information Criteria and Statistical Modeling*. Springer, 2008.

[82] S. D. Kost, Y.K. Dewaraja, R. G. Abramson, and M. G. Stabin. VIDA: A Voxel-Based Dosimetry Method for Targeted Radionuclide Therapy Using Geant4. *Cancer Biotherapy and Radiopharmaceuticals*, 30(1):16–26, 2015. PMID: 25594357.

[83] M.H. Kulke, L.L. Siu, J.E. Tepper, G. Fisher, D. Jaffe, D.G. Haller, L.M. Ellis, J.K. Benedetti, E.K. Bergsland, T.J. Hobday, et al. Future directions in the treatment of neuroendocrine tumors: consensus report of the National Cancer Institute Neuroendocrine Tumor clinical trials planning meeting. *Journal of Clinical Oncology*, 29(7):934, 2011.

[84] N. Lanconelli, M. Pacilio, S. Lo Meo, F. Botta, A. Di Dia, L.A. Torres Aroche, M.A. Coca Pérez, and M. Cremonesi. A free database of radionuclide voxel S values for the dosimetry of nonuniform activity distributions. *Physics in Medicine and Biology*, 57:517–533, 2012.

[85] M. Lassmann, C. Chiesa, G. Flux, and M. Bardiès. EANM Dosimetry Committee guidance document:good practice of clinical dosimetry reporting. *European Journal of Nuclear Medicine and Molecular Imaging*, 38(1):192–200, 2011.

[86] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[87] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1:541–551, 1989.

[88] A. Linderhed. 2-D empirical mode decompositions in the spirit of image compression. In *Wavelet and Independent Component Analysis Applications IX*, pages 1–8. Proceedings of SPIE Vol. 4738, 2002.

[89] B.V. Lingme. *A Guide to Microsoft Exel 2013 for Scientists and Engineers*. Elsevier, 2016.

[90] A. Liu, L.E. Williams, and G. Lopatin. A radionuclide therapy treatment planning and dose estimation system. *Journal of Nuclear Medicine*, 40:1151–1153, 1999.

[91] Z. Liu and S. Peng. Boundary processing of bidimensional EMD using texture synthesis. *IEEE Signal Processing Letters*, 12:33–36, 2005.

[92] Z. Liu, H. Wang, and S. Peng. Texture classification through directional empirical mode decomposition. In *Proc. 17th IEEE International Conference on Pattern Recognition (ICPR '04)*, pages 803–806, 2004.

[93] M. Ljungberg and K. Sjögreen Gleisner. Personalized Dosimetry for Radionuclide Therapy Using Molecular Imaging Tools. *Biomedicines*, 4:25, 2016.

[94] R. Loevinger, T.F. Budinger, and E.E. Watson. *MIRD Primer for Absorbed Dose Calculations*. The Society of Nuclear Medicine, Revised ed., New York, NY, 1991.

[95] Bishop Ch. M. *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.

[96] D.J.C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[97] L. Maffioli, L. Florimonte, D.C. Costa, J. Correia Castanheira, C. Grana, M. Luster, L. Bodei, and M. Chinol. New radiopharmaceutical agents for the treatment of castration-resistant prostate cancer. *QJ Nuclear Medicine and Molecular Imaging*, 59(4):420–438, 2015.

[98] S. Marcatili. Multi-scale dosimetry for targeted radionuclide therapy optimisation, 2015.

[99] S. Marcatili, C. Pettinato, S. Daniels, G. Lewis, P. Edwards, S. Fanti, and E. Spezi. Development and validation of RAYDOSE: a Geant4-based application for molecular radiotherapy. *Physics in Medicine and Biology*, 58(8):2491, 2013.

[100] K.M. Maryam, K.A. Alireza, G. Parham, and Z. Habib. Evaluating the Application of Tissue-Specific Dose Kernels instead of Water Dose Kernels in Internal Dosimetry: A Monte Carlo study. *Cancer Biotherapy and Radiopharmaceuticals*, 31(10):367–379, 2016.

[101] A. Maze, J. Le Cloirec, R. Collorec, Y. Bizais, P. Briandet, and P. Bourguet. Iterative reconstruction methods for nonuniform attenuation distribution in SPECT. *Journal of Nuclear Medicine*, 34:1204–1209, 1993.

[102] K. Min-Sung, E. Rodriguez-Marek, and T.R. Fischer. A new two dimensional empirical mode decomposition for images using inpainting. In *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, pages 13–16, 2010.

[103] A.R. Moattari, L.J. Deftos, and A.I. Vinik. Effects of sandostatin on plasma chromogranin-A levels in neuroendocrine tumors. *The Journal of Clinical Endocrinology & Metabolism*, 69(4):902–905, 1989.

[104] I.M. Modlin, B. I. Gustafsson, S. F. Moss, M. Pavel, A. V. Tsolakis, and M. Kidd. Chromogranin A—biological function and clinical utility in neuro endocrine tumor disease. *Annals of surgical oncology*, 17(9):2427–2443, 2010.

[105] V. Nair and G. E. Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proc. 27th International Conference Machine Learning*, pages 1–8, Haifa, Israel, 2010.

[106] NCPR164. Uncertainties in Internal Radiation Dose Assessment, 2009.

[107] M. Nickel, S.E. Strand, O. Linden, K. Wingardh, J. Tennwall, and K.S. Gleisner. Development and evaluation of a pharmacokinetic model for prediction of radioimmunotherapy based on pretherapy data. *Cancer biotherapy & radiopharmaceuticals*, 24:111–122, 2009.

[108] J.C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and Bunel P. Image Analysis by Bidimensional Empirical Mode Decomposition. *Image and Vision Computing*, 21(12):1019–1026, 2003.

[109] J.C. Nunes and E. Deléchelle. Empirical mode decomposition: Applications on signal and image processing. *Advances in Adaptive Data Analysis*, 1:125–75, 2009.

[110] J.C. Nunes, S. Guyot, and E. Deléchelle. Texture analysis based on local analysis of the bidimensional empirical mode decomposition. *Machine Vision and Applications*, 16:177–188, 2005.

[111] A. O'Hagan and J. Forster. *Kendall's Advanced Theory od Statistics, Vol. 2B: Bayesian Inference*. Arnold Publishers, 1999.

[112] S. Okamoto, A. Thieme, J. Allmann, C. D'Alessandria, T. Maurer, M. Retz, R. Tauber, M. M. Heck, H.J. Wester, N. Tamaki, et al. Radiation dosimetry for 177Lu-PSMA I&T in metastatic castration-resistant prostate cancer: absorbed dose in normal organs and tumor lesions. *Journal of Nuclear Medicine*, 58(3):445–450, 2017.

[113] M. Pacilio, E. Amato, N. Lanconelli, C. Basile, L. Torres, F. Botta, M. Ferrari, N. Diaz, M. Perez, M. Fernández, et al. Differences in 3D dose distributions due to calculation method of voxel S-values and the influence of image blurring in SPECT. *Physics in medicine and biology*, 60(5):1945, 2015.

[114] M. Pacilio, N. Lanconelli, S. Lo Meo, M. Betti, L. Montani, L. Torres Aroche, and C. Pérez. Differences among Monte Carlo codes in the calculations of voxel S values for radionuclide targeted therapy and analysis of their impact on absorbed dose evaluations. *Medical physics*, 36(5):1543–1552, 2009.

[115] P. Papadimitroulas, G. Loudos, G. Nikiforidis, and G. Kagadis. A dose point kernel database using GATE Monte Carlo simulation toolkit for nuclear medicine applications: Comparison with other Monte Carlo codes. *Medical physics*, 39(8):5238–5247, 2012.

[116] C. Parker, S. Nilsson, D. Heinrich, S. I. Helle, J.M. O'sullivan, S.D. Fosså, A. Chodacki, P. Wiechno, J. Logue, M. Seke, et al. Alpha emitter radium-223 and survival in metastatic prostate cancer. *New England Journal of Medicine*, 369(3):213–223, 2013.

[117] S. Pauwels, R. Barone, and St. et al. Walrand. Practical dosimetry of peptide receptor radionuclide therapy with ^{90}Y-labeled somatostatin analogs. *Journal of Nuclear Medicine*, 46:92S–98S, 2005.

[118] K.I. Penny. Appropriate critical values when testing for a single multivariate outlier by using the mahalanobis distance. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 45(1):73–81, 1996.

[119] P. Pérez, F. Botta, G. Pedroli, and M. Valente. *Dosimetry for beta-emitter radionuclides by means of Monte Carlo simulations*. INTECH Open Access Publisher, 2011.

[120] S. Perner, M.D. Hofer, R. Kim, R. B. Shah, H. Li, P. Möller, R.E. Hautmann, J.E. Gschwend, R. Kuefer, and M. A Rubin. Prostate-specific membrane antigen expression as a predictor of prostate cancer progression. *Human pathology*, 38(5):696–701, 2007.

[121] N. Petoussi-Henss, W. Bolch, M. Zankl, G. Sgouros, and B. Wessels. Patient-specific scaling of reference S-values for cross-organ radionuclide S-values: what is appropriate? *Radiation protection dosimetry*, 127(1-4):192–196, 2007.

[122] G.R. Pond, G. Sonpavde, R. De Wit, M.A. Eisenberger, I.F. Tannock, and A.J. Armstrong. The prognostic importance of metastatic site in men with metastatic castration-resistant prostate cancer. *European urology*, 65(1):3–6, 2014.

[123] A. Prideaux, H. Song, R. Hobbs, B. He, E. Frey, P. Ladenson, R. Wahl, and Sgouros G. 3D radiobiologic dosimetry: application of radiobiologic modelling to patient-specific 3D imaging-based internal dosimetry. *Journal of Nuclear Medicine*, 48:1008–1016, 2007.

[124] K. Rahbar, H. Ahmadzadehfar, C. Kratochwil, U. Haberkorn, M. Schäfers, M. Essler, R.P. Baum, H.R. Kulkarni, M. Schmidt, A. Drzezga, et al. German multicenter study investigating 177Lu-PSMA-617 radioligand therapy in advanced prostate cancer patients. *Journal of Nuclear Medicine*, 58(1):85–90, 2017.

[125] D. Reiner, M. Blaickner, and F. Rattay. Discrete beta dose kernel matrices for nuclides applied in targeted radionuclide therapy (TRT) calculated with MCNP5. *Medical physics*, 36(11):4890–4896, 2009.

[126] G. Rilling, P. F1andrin, P. Goncalves, and J.M. Lilly. Bivariate empirical mode decomposition. *IEEE Signal Processing Letter*, 14:936–939, 2007.

[127] A. Rinke, H.H. Muller, C. Schade-Brittinger, K.J. Klose, P. Barth, M. Wied, C. Mayer, B. Aminossadati, U.F. Pape, M. Blaker, et al. Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide lar in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: a report from the PROMID Study Group. *Journal of Clinical Oncology*, 27(28):4656–4663, 2009.

[128] P. Ritt, J. Hornegger, and T. Kuwert. Technik und physikalische Aspekte der SPECT/CT. *Der Nuklearmediziner*, 34(01):9–20, 2011.

[129] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional Networks for Biomedical Image Segmentation. *Clinical Orthopaedics and Related Research*, abs/1505.04597, 2015.

[130] S. Ruder. An overview of gradient descent optimization algorithms, 2017.

[131] D. Salas-González, J.M. Górriz, J. Ramírez, M. Schloegl, E.W. Lang, and A. Ortiz. Parameterization of the distribution of white and grey matter in MRI using the α-stable distribution. *Computers in biology and medicine*, 43(5):559–567, 2013.

[132] J.C. Sanders, T. Kuwert, J. Hornegger, and Ph. Ritt. Quantitative SPECT/CT imaging of 177-Lu with In Vivo Validation in Patients Undergoing Peptide Receptor Radionuclide Therapy. *Molecular Imaging and Biology*, 2014.

[133] M. Sandström, U. Garske, D. Granberg, A. Sundin, and H. Lundqvist. Individualized dosimetry in patients undergoing therapy with 177 Lu-DOTA-D-Phe 1-Tyr 3-octreotate. *European journal of nuclear medicine and molecular imaging*, 37(2):212–225, 2010.

[134] D.T. Sandwell. Biharmonic spline interpolation of Geos-3 and Seasat altimeter data. *Geophysics Research Lett.*, 14(2):139–142, 1987.

[135] D. Sarrut, J.N. Badel, A. Halty, G. Garin, D. Perol, P. Cassier, J.Y. Blay, D. Kryza, and A.L. Giraudet. 3D absorbed dose distribution estimated by

Monte Carlo simulation in radionuclide therapy with a monoclonal antibody targeting synovial sarcoma. *European Journal of Nuclear Medicine and Molecular Imaging Physics*, 4(1):6, Jan 2017.

[136] D. Sarrut, A. Halty, J.N. Badel, L. Ferrer, and M. Bardis. Voxel-based multimodel fitting method for modeling time activity curves in SPECT images. *Medical Physics*, 44(12):6280–6288, 2017.

[137] I. Scarinci, M. Valente, and Pérez. Dose Point Kernel calculation and modelling with nuclear medicine dosimetry purposes.

[138] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[139] W. Schneider, T. Bortfeld, and W. Schlegel. Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions. *Physics in Medicine and Biology*, 45(2):459, 2000.

[140] C. Schuchardt, H. R. Kulkarni, V. Prasad, C. Zachert, D. Müller, and R.P. Baum. The Bad Berka dose protocol: comparative results of dosimetry in peptide receptor radionuclide therapy using 177 Lu-DOTATATE, 177 Lu-DOTANOC, and 177 Lu-DOTATOC. In *Theranostics, Gallium-68, and Other Radionuclides*, pages 519–536. Springer, 2013.

[141] G. Sgouros, E. Frey, R. Wahl, B. He, A. Prideaux, and R. Hobbs. Three-dimensional imaging-based radiobiological dosimetry. In *Seminars in nuclear medicine*, volume 38, pages 321–334. Elsevier, 2008.

[142] G. Sgouros, K. Kolbert, A. Sheikh, K. Pentlow, E. Mun, A. Barth, R. Robbins, and S. Larson. Patient-specific dosimetry for $131I$ thyroid cancer therapy using $124I$ PET and 3-dimensional-internal dosimetry (3D-ID) software. *Journal of Nuclear Medicine*, 45(8):1366–1372, 2004.

[143] G. Sgouros, K. S. Kolbert, A. Sheikh, K. S. Pentlow, et al. Patient-specific dosimetry for 131-I thyroid cancer therapy using 124-I PET and 3-dimensional-internal dosimetry (3D-ID) software. *The Journal of Nuclear Medicine*, 45(8):1366, 2004.

[144] G. Sgouros, S. Squeri, and A.M. et al. Ballangrud. Patient-specific, 3-dimensional dosimetry in non-Hodgkin's lymphoma patients treated with ^{131}I-anti-B1 antibody: assessment of tumor dose-response. *Journal of Nuclear Medicine*, 44:260–268, 2003.

[145] J.A. Siegel, S.R. Thomas, J.B. Stubbs, M.G. Stabin, M.T. Hays, K.F. Koral, J.S. Robertson, R.W. Howell, D. R. Wessels, B. W .and Fisher, et al. MIRD pamphlet no. 16: techniques for quantitative radiopharmaceutical biodistribution data acquisition and analysis for use in human radiation dose estimates. *Journal of Nuclear Medicine*, 40(2):37S, 1999.

[146] S. S. Sridhar, S. J. Freedland, M. E. Gleave, C. Higano, P. Mulders, C. Parker, O. Sartor, and F. Saad. Castration-resistant prostate cancer: from new pathophysiology to new treatment. *European urology*, 65(2):289–299, 2014.

[147] M. Stabin. The case for patient-specific dosimetry in radionuclide therapy. *Cancer biotherapy & radiopharmaceuticals*, 23(3):273–284, 2008.

[148] M. Stabin and X.G. Xu. Basic Principles in the Radiation Dosimetry of Nuclear Medicine. *Seminars in Nuclear Medicine*, 44:162–171, 2014.

[149] M.G. Stabin, R.B. Sparks, and E. Crowe. OLINDA/EXM: the second-generation personal computer software for internal dose assessment in nuclear medicine. *Journal of Nuclear Medicine*, 46:1023–1027, 2005.

[150] M.G. Stabin, M. Tagesson, S.R. Thomas, M. Ljungberg, and S.E. Strand. Radiation dosimetry in nuclear medicine. *Applied Radiation and Isotopes*, 50:73–87, 1999.

[151] M. Stridsberg, K. Öberg, Q. Li, U. Engström, and G. Lundqvist. Measurements of chromogranin A, chromogranin B (secretogranin I), chromogranin C (secretogranin II) and pancreastatin in plasma and urine from patients with carcinoid tumours and endocrine pancreatic tumours. *Journal of Endocrinology*, 144(1):49–59, 1995.

[152] L. Strifari, E. Menghi, M.D. Andrea, and M. Benassi. Monte Carlo dose voxel kernel calculations of beta-emitting and Auger-emitting radionuclides for internal dosimetry: A comparison between EGSnrcMP and EGS4. *Medical Physics*, 33(9):3383–3389, 2006.

[153] J. Strosberg, G. El-Haddad, E. Wolin, A. Hendifar, J. Yao, B. Chasen, E. Mittra, P. L. Kunz, M. H. Kulke, H. Jacene, et al. Phase 3 trial of 177Lu-DOTATATE for midgut neuroendocrine tumors. *New England Journal of Medicine*, 376(2):125–135, 2017.

[154] C. J. Sweeney, Y.H. Chen, M. Carducci, G. Liu, M. Jarrard, D.F .and Eisenberger, Y.N. Wong, N. Hahn, M. Kohli, M. M. Cooney, et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer. *New England Journal of Medicine*, 373(8):737–746, 2015.

[155] S.P. Thang, M. S. Lung, G. Kong, M. S. Hofman, J. Callahan, M. Michael, and R. J. Hicks. Peptide receptor radionuclide therapy (PRRT) in European Neuroendocrine Tumour Society (ENETS) grade 3 (G3) neuroendocrine neoplasia (NEN)-a single-institution retrospective analysis. *European journal of nuclear medicine and molecular imaging*, 45(2):262–277, 2018.

[156] P. Thapa, R. Ranade, V. Ostwal, S.V. Shrikhande, M. Goel, and S. Basu. Performance of 177Lu-DOTATATE-based peptide receptor radionuclide therapy in metastatic gastroenteropancreatic neuroendocrine tumor: a multiparametric response evaluation correlating with primary tumor site, tumor proliferation index, and dual tracer imaging characteristics. *Nuclear medicine communications*, 37(10):1030–1037, 2016.

[157] A. Traino, S. Marcatili, C. Avigo, M. Sollini, P. Erba, and G. Mariani. Dosimetry for nonuniform activity distributions: A method for the calculation of 3D absorbed-dose distribution without the use of voxel S-values, point kernels, or Monte Carlo simulations. *Medical physics*, 40(4), 2013.

[158] M. Van Essen, E. P. Krenning, M. De Jong, R. Valkema, and D.J. Kwekkeboom. Peptide receptor radionuclide therapy with radiolabelled somatostatin analogues in patients with somatostatin receptor positive tumours. *Acta Oncologica*, 46(6):723–734, 2007.

[159] P. Vicini, A. B. Brill, and A. Stabin, M.and Rescigno. Kinetic modeling in support of radionuclide dose assessment. *Seminars in nuclear medicine*, 38(5):335–346, 2008.

[160] D. Villoing, S. Marcatili, M.P. Garcia, and M. Bardiès. Internal dosimetry with Monte Carlo code GATE: validation using ICRP/ICRU female reference computational model. *Physics in Medicine and Biology*, 62:1885–1904, 2017.

[161] J. A. Violet, P. Jackson, J. Ferdinandus, S. Sandhu, T. Akhurst, A. Iravani, G. Kong, A.R. Kumar, S. P. Thang, P. Eu, et al. Dosimetry of Lu-177 PSMA-617 in metastatic castration-resistant prostate cancer: correlations between pre-therapeutic imaging and "whole body" tumor dosimetry with treatment outcomes. *Journal of Nuclear Medicine*, pages jnumed–118, 2018.

[162] C. Wehrmann, S. Senftleben, C. Zachert, D. Müller, and R.P. Baum. Results of individual patient dosimetry in peptide receptor radionuclide therapy with 177Lu DOTA-TATE and 177Lu DOTA-NOC. *Cancer biotherapy & radiopharmaceuticals*, 22(3):406–416, 2007.

[163] P. Wessel. A general-purpose Green's function-based interpolator. *Computers and Geosciences*, 35(6):1247 – 1254, 2009.

[164] P. Wessel and D. Bercovici. Interpolation with Splines in Tension: A Green's Function Approach. *Mathematical Geology*, 30(1):77–93, 1998.

[165] S.J. Wilderman, A.M. Avram, J. Kritzman, R. Ackerman, and Y.K. Dewaraja. Dosimetry in 131I internal emitter therapy using voxel dependent integrated time-activities derived from multiple, registered SPECT and CT images. In *Nuclear Science Symposium Conference Record, 2006. IEEE*, volume 6, pages 3492–3496. IEEE, 2006.

[166] Z. Wu, N.E. Huang, and X. Chen. The Multidimensional Ensemble Empirical Mode Decomposition Method. *Adv. Adaptive Data Analysis*, 1:339–372, 2009.

[167] C.Z. Xiong, J.Y. Xu, J.C. Zou, and D.X. Qi. Texture classification based on EMD and FFT. *Journal of Zhejiang University - Science A*, 7:1516–1521, 2006. 10.1631/jzus.2006.A1516.

[168] M.P. Yadav, S. Ballal, M. Tripathi, N.A. Damle, R.K. Sahoo, A. Seth, and C. Bal. 177 Lu-DKFZ--617 therapy in metastatic castration resistant prostate cancer: safety, efficacy, and quality of life assessment. *European journal of nuclear medicine and molecular imaging*, 44(1):81–91, 2017.

[169] J. C. Yao, M. Hassan, A. Phan, C. Dagohoy, C. Leary, J. E. Mares, E. K. Abdalla, J. B. Fleming, J.N. Vauthey, A. Rashid, et al. One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the united states. *Journal of clinical oncology*, 26(18):3063–3072, 2008.

[170] J.C. Yao, N. Fazio, S. Singh, R. Buzzoni, C. Carnaghi, E. Wolin, J. Tomasek, M. Raderer, H. Lahner, M. Voi, et al. Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): a randomised, placebo-controlled, phase 3 study. *The Lancet*, 387(10022):968–977, 2016.

[171] H. Yoriyaz, M. Stabin, and Dos Santos A. Monte Carlo MCNP-4B-based absorbed dose distribution estimates for patient-specific dosimetry. *Journal of Nuclear Medicine*, 42:662–669, 2001.

[172] K. Yun, A. Huyen, and T. Lu. Deep Neural Networks for Pattern Recognition, 2018.

[173] P.A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J.C. Gee, and G. Gerig. User-guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. *Neuroimage*, 31(3):1116–1128, 2006.

Acknowledgements

I want to thank the following persons:

- Prof. Dr.-Ing. Bernd Ludwig for the brilliant supervision,
- Prof. Dr. Elmar W. Lang for the excellent support and the helpful advices,
- Prof. Dr. Torsten Kuwert, who has given me the opportunity to conduct this thesis in his group,
- Dr. Christian Schmidkonz for the medical guidance,
- All members of the AG Lang,
- My family and friends for their support during my study.