



PAPER

A deep learning approach to radiation dose estimation

RECEIVED
24 August 2019REVISED
29 November 2019ACCEPTED FOR PUBLICATION
27 December 2019PUBLISHED
4 February 2020Th I Götz^{1,2,3,4} , C Schmidkonz¹, S Chen³, S Al-Baddai², T Kuwert¹ and E W Lang² ¹ Clinic of Nuclear Medicine, University Hospital Erlangen, 91054 Erlangen, Germany² CIML Group, Biophysics, University of Regensburg, 93040 Regensburg, Germany³ Pattern Recognition Lab, University of Erlangen-Nürnberg, 91058 Erlangen, Germany⁴ Author to whom any correspondence may be addressed.E-mail: theresa.goetz@biologie.uni-regensburg.de**Keywords:** Monte Carlo simulation, dosimetry, Lu-therapy, neural network**Abstract**

Currently methods for predicting absorbed dose after administering a radiopharmaceutical are rather crude in daily clinical practice. Most importantly, individual tissue density distributions as well as local variations of the concentration of the radiopharmaceutical are commonly neglected. The current study proposes machine learning techniques like Green's function-based empirical mode decomposition and deep learning methods on U-net architectures in conjunction with soft tissue kernel Monte Carlo (MC) simulations to overcome current limitations in precision and reliability of dose estimations for clinical dosimetric applications. We present a hybrid method (DNN-EMD) based on deep neural networks (DNN) in combination with empirical mode decomposition (EMD) techniques. The algorithm receives x-ray computed tomography (CT) tissue density maps and dose maps, estimated according to the MIRD protocol, i.e. employing whole organ S-values and related time-integrated activities (TIAs), and from measured SPECT distributions of ¹⁷⁷Lu radionuclei, and learns to predict individual absorbed dose distributions. In a second step, density maps are replaced by their intrinsic modes as deduced from an EMD analysis. The system is trained using individual full MC simulation results as reference. Data from a patient cohort of 26 subjects are reported in this study. The proposed methods were validated employing a leave-one-out cross-validation technique. Deviations of estimated dose from corresponding MC results corroborate a superior performance of the newly proposed hybrid DNN-EMD method compared to its related MIRD DVK dose calculation. Not only are the mean deviations much smaller with the new method, but also the related variances are much reduced. If intrinsic modes of the tissue density maps are input to the algorithm, variances become even further reduced though the mean deviations are less affected. The newly proposed hybrid DNN-EMD method for individualized radiation dose prediction outperforms the MIRD DVK dose calculation method. It is fast enough to be of use in daily clinical practice.

1. Introduction

Current clinical practice in radiotherapy often still relies on rather crude absorbed dose prediction methods because of high computational complexity and time-consuming data handling of more sophisticated methods. The standard dose estimation method in current clinical practice mostly relies on the simplest and least precise approach by estimating, according to the Medical Internal Radiation Dose Committee (MIRD) protocol (Bolch *et al* 1999), time-integrated activities (TIAs) and dose kernels, which the MIRD protocol calls S-values. While the former has been the subject of a recent study (Götz *et al* 2019), the latter combine mass density information deduced from a standard body phantom (Petoussi-Henss *et al* 2007) with either continuous dose point kernels (DPK) (Pérez *et al* 2011) or freely scalable discrete dose voxel kernels (DVK) (Pacilio *et al* 2015). Both DPK and DVK methods are subject to substantial uncertainties (Götz *et al* 2019b) and employ Monte Carlo simulations (Papadimitroulas *et al* 2012) performed on homogeneous soft tissue kernels with a limited number of radioisotopes. All these dose estimation procedures still have the drawback of not properly taking into account individual mass density and radioactivity distributions. But such personalized information is available from

modern hybrid imaging technologies such as SPECT/CT or PET/CT (Cal-Gonzalez *et al* 2018). However, an increasing number of therapies with radiopharmaceuticals renders a patient-specific dosimetry of fundamental importance (Botta *et al* 2011, Lou *et al* 2019).

Early attempts to predict radiation dose with neural networks still relied on fully connected networks, which, however, tend to overfit data and badly generalize on big data settings (Kalantzis *et al* 2011). Given the recent success of deep convolutional neural networks (CNNs) on image processing (Chen *et al* 2017, Kearney *et al* 2018c, Park *et al* 2018) and image segmentation tasks (Jackson *et al* 2018) employing U-Net or V-Net architectures (Ronneberger *et al* 2015), Kearney *et al* (2018a), very recently deep learning techniques and deep neural network architectures were employed in external radiotherapy dosimetry. Nguyen *et al* (2017) were the first to employ an U-Net architecture in a radiation dose prediction feasibility study. Shortly afterwards (Kearney *et al* 2018b) introduced DoseNet, a 3D convolutional volumetric dose prediction neural network. DoseNet used an U-Net architecture amended with residual blocks which reduced network redundancy. It was trained and tested on a cohort of 150 prostate cancer patients and the dosimetric quality of its predictions was evaluated on various clinically relevant dosimetric parameters. While the current study was under review for publication, an internal radiation dose prediction study employing deep learning and convolutional neural networks was published by Lee *et al* (2019). The network received 3D information about radioactivity distribution from positron emission tomography (PET) images and 3D tissue density information from x-ray computed tomography (CT) images. The CNN was trained to predict 3D dose rate maps, which were evaluated against reference Monte Carlo simulations.

Though the computational complexity and computational load of full whole body MC simulations is prohibitive for daily clinical practice, they still will be used in this study to provide reference data for the development of approximate dose estimation methods. The latter should be both reliable and fast enough for clinical use. Following, we apply machine learning techniques like Green's function-based empirical mode decomposition (AL-Baddai *et al* 2016) and deep learning methods on U-net architectures (Ronneberger *et al* 2015) in conjunction with soft tissue kernel MC simulations to overcome current limitations in precision and reliability of dose estimations for clinical dosimetric applications.

2. Methods

2.1. Patient data

For network training, data from 26 patients was available. The patients, which all obtained a ^{177}Lu -PSMA therapy, consisted of 3 female and 23 male patients with an average age of 63.8 ± 10 years (Götz *et al* 2019). For each patient, one CT image and SPECT images at four different time points were available.

2.1.1. SPECT

From every patient four SPECT images were acquired at $t = 4, 24, 48, 72$ h after administering the radiopharmaceutical. At the second time point, i.e. after 24 h, data was recorded in dual mode on a hybrid SPECT/CT instrument and manually co-registered by a medical expert. The CT image was further used for attenuation correction of the SPECT image. Image acquisition followed the standard quantitative ^{177}Lu —protocol as described in Sanders *et al* (2015), hence it is only summarized as follows:

- SPECT using medium energy collimators, 3° angular sampling, $t_d = 15$ min total dwell time
- Iterative OSCG (see below) reconstruction of the 208 keV photo-peak data with 24 iterations, 1 subset, matrix 256×256
- Point-spread-function modeling in reconstruction
- Triple energy window based scatter correction
- CT-based attenuation correction
- No post-reconstruction smoothing.

Image reconstruction was performed with the ordered-subset conjugate-gradient (OSCG) algorithm, which outputs data in the same coordinate system as the CT data. The voxel size was $(4.79 \text{ mm})^3$ and the reconstructed SPECT images were cropped to 82^3 voxels, removing large portions of the reconstructed volume which predominantly contained air.

2.1.2. X-ray CT

An additional low dose CT has been recorded for SPECT attenuation correction. It covered the same field-of-view as the SPECT and was acquired and reconstructed using the following parameters:

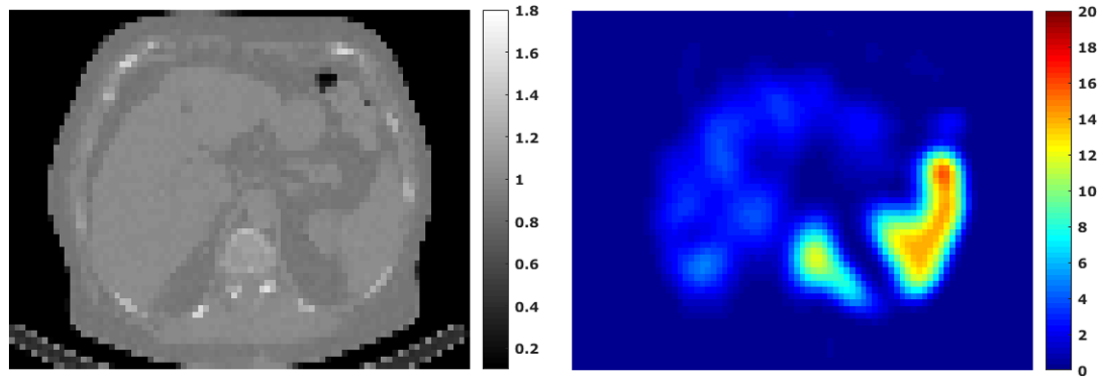


Figure 1. On the left side a density map in g/cm^3 , corresponding to one slice of the 3D CT image, is displayed, and on the right side the corresponding activity map in Bq is shown for a prostate cancer patient.

- Slice collimation of 2×5 mm, pitch of 1.8, time per rotation of $t = 0.8$ s, tube voltage of $V = 130$ kVp, effective tube current of $J = 30$ mAs
- Filtered-back-projection reconstruction with B08s and B41s Kernels, 512×512 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 (VOIs).

The mass density distribution from each CT image slice was resampled to a *density map* with voxel size of $(4.89 \text{ mm})^3$, which matched the spatial resolution of the SPECT images. Photon attenuation coefficients are expressed in the usual hounsfield units (HU) and are due to photon—matter interactions. Because of the latter, attenuation coefficients are strongly correlated with electron densities, which themselves are correlated to mass densities. Thus CT images can be used to derive mass densities, usually by piecewise linear relations with different slopes for $HU \leq 0$ and $HU > 0$. However, a system specific calibration with dedicated phantoms has not been done, rather mass density was computed according to

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

whereby mass density ρ is expressed in CGS units g/cm^3 . The four SPECT images were integrated over time voxelwise to obtain the corresponding *activity maps* (see figure 1) as is described in detail in Götz *et al* (2019, 2019b). In the latter study, a particle filter (PF) has been applied for de-noising the time-activity-curves (TACs) of the patient cohort. The TACs were obtained from fully-quantitative, serial SPECT(/CT) data, acquired at 4 h, 24 h, 48 h, 72 h p.i. The model used in the PF was a mono-exponential decay and its free parameters were determined based on objective criteria. The time-integrated activities (TIA), computed after Bayesian filtering with a PF, were compared to the results of a mono-exponential fit to individual voxel TACs (simple fit) in several volumes of interest (kidneys, spleen, tumours). Additionally, an organ-averaged TIA was derived from whole-organ mean TACs resulting from mean activation time series and subsequent curve-fitting. This organ-averaged TIA was also compared to the whole-organ TIAs obtained from summation of the voxel-wise TIAs from both, the PF-approach and the simple fit approach (Götz *et al* 2019).

2.2. Monte Carlo simulations

To provide proper reference data, the *activity map* and the corresponding mass *density map* were input to the software package GAMOS to perform a full whole-body Monte Carlo (MC) simulation yielding the related spatial distribution of absorbed energy dose, henceforth called *dose map*. In figure 2, the energy dose distribution corresponding to figure 1 is displayed exemplarily.

Thus, for generating reference data, in this study a full MC simulation of radiation transport was carried out for every patient. For this, the GEANT4-based GAMOS 5.0.0 toolkit was used (Arce *et al* 2014). The procedure is briefly described in the following:

- A voxelized geometry was generated based on pre-defined tissue classes and mass-density of each voxel. The size of the latter was $4.8 \times 4.8 \times 4.8 \text{ mm}^3$.
- The decay modes of ^{177}Lu and emitted gamma and beta energy spectra were defined according to Arce *et al* (2014).

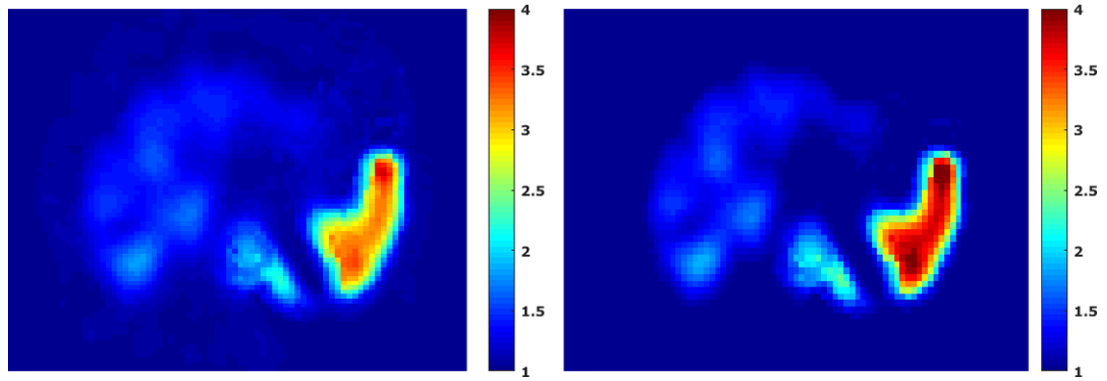


Figure 2. On the left side one slice of the dose map calculated via a whole body MC simulation is illustrated, and on the right side the dose distribution determined with the MIRD standard protocol is shown. The colorbar is in Gy.

- The relative spatial distribution of radioactive decays, i. e., the number of nuclear disintegrations per voxel divided by the total number of disintegrations, was taken from the activity distribution as deduced from SPECT images.
- Deposited energy was scored per voxel in units of Gray per decay.

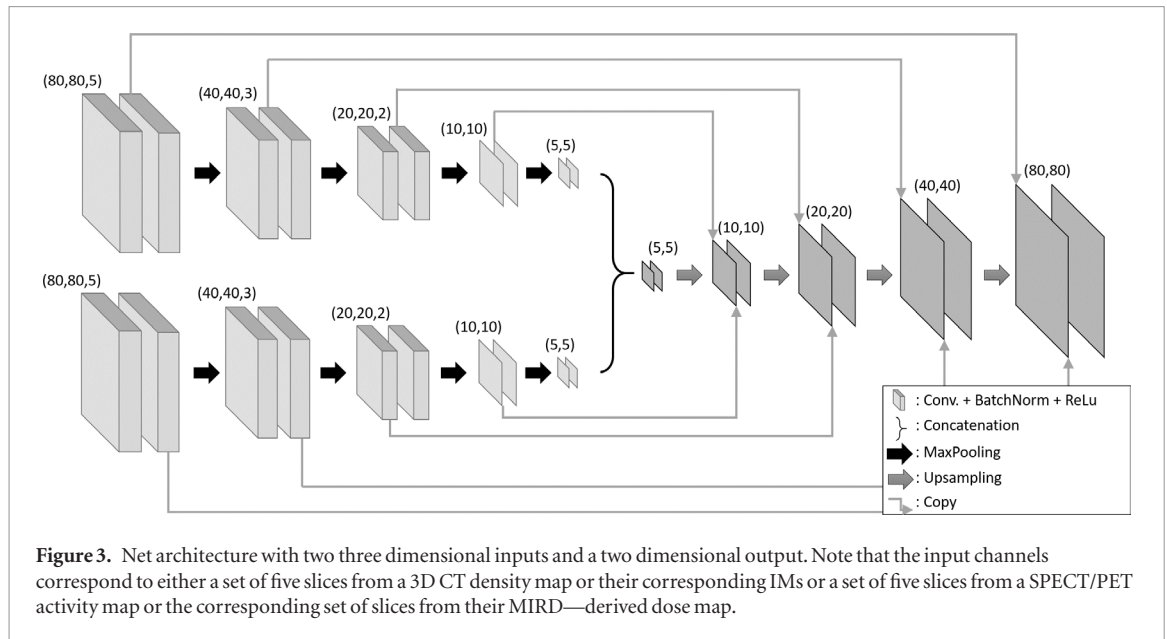
2.3. MIRD DVK-based dose calculation

To follow the MIRD protocol, DVKs have to be estimated via MC simulations of defined kernels with fixed size, defined homogeneous tissue density and given radioisotope. In this study the DVK comprised 9^3 voxels (size $(4.8 \text{ mm})^3$) which consisted of soft tissue ($\rho_m = 1043 \text{ kg m}^{-3}$) and had a ^{177}Lu -isotope placed in the center voxel. Thus the isotope was placed with equal probability anywhere in the center voxel of a $9 \times 9 \times 9$ cube of voxels consisting of a specific tissue-type. For every patient, in total 10^9 nuclear disintegrations of ^{177}Lu -isotopes were simulated using GAMOS on a high-performance cluster equipped with 176 nodes. This number of radioactive decays assured a confidence level γ higher than 99% per voxel, where $\gamma = 100\% \cdot (1 - \alpha)$ and α measures the relative statistical uncertainty. Via a MC simulation, the energy dose per ^{177}Lu -decay was calculated for each voxel of the kernel. The resulting DVK was convolved with the activity map to obtain the dose map. This method yields reasonable dose estimates for soft tissue, but, unfortunately, it suffers from large density differences in bone and lung tissue relative to soft tissue. Tissue classes and related density ranges have been determined as described in Schneider *et al* (2000) and Götz *et al* (2019a). Both, the dose maps resulting from a full-body MC simulation and this MIRD DVK dose calculation are illustrated for the slice of figures 1 in 2.

2.4. A hybrid DNN-EMD method

This study proposes to replace the computationally expensive MC simulations by a deep neural network (DNN) trained with the data set described above. A modified U-net architecture (Ronneberger *et al* 2015) was employed with both a *density map* $\mathbf{M}(\mathbf{r}_T)$, as deduced from x-ray CT images, and an *activity map* $\tilde{\mathbf{A}}(\mathbf{r}_S)$, as deduced from SPECT images, as inputs and a corresponding *dose map* $\mathbf{D}(\mathbf{r}_T)$ as output. Here $\mathbf{r}_T, \mathbf{r}_S$ denote voxel locations in the target (T) regions and the source (S) regions, respectively. More specifically, for the current study, we replace the *activity map* $\tilde{\mathbf{A}}$ by its related dose map $\tilde{\mathbf{D}}$ as estimated with the standard MIRD protocol. We also compare this new method to a further variant, where we replace the *density map* $\mathbf{M}(\mathbf{r}_T)$ by its intrinsic spatial modes (IMs) extracted via a Green's function based bi-dimensional ensemble empirical mode decomposition (GiT-BEEMD). All predicted dose maps are finally compared to reference MC dose distributions. The comparison is confined to the following organs: kidneys, spleen and all tracer-positive metastases.

For training the modified U-net, the 3D SPECT/CT imaging data was partitioned into packs of 5 adjacent slices each. This seemed appropriate in view of the short range of electrons emitted from ^{177}Lu in condensed matter, which is less than 3 mm. The slices were further resampled to a size of $(n_x = 80) \times (n_y = 80)$ voxels. Thus there will be two 3D input arrays of size $(n_x = 80) \times (n_y = 80) \times (n_z = 5)$ and one 2D output of size $(n_x = 80) \times (n_y = 80)$. All input arrays were normalized to the respective maximal value of the entire data set. In figure 3, the architecture of the modified U-net is illustrated. The encoding paths consist of five blocks, which comprise of the following sequence of operations that were repeated twice: a convolution operation with m different convolution kernels (henceforth called filters), each of size $k_x \times k_y$ or $k_x \times k_y \times k_z$, followed by a batch normalization and a leaky rectified linear unit (ReLU) activation. The convolution kernel was sliding over the input array with a stride of one, while zero padding was applied to preserve the size of the array. The resulting feature map was down-sampled employing a max-pooling kernel of size $(p_x = 2) \times (p_y = 2)$ to each slice while



skipping the two outermost slices along the z -direction until only the central slice remained. Filter size $k \times k$ and number of filters m were varied in steps of two, i.e. $k = 5, 7, 9$ and $m = 4, 6, 8$. As loss function, the mean squared reconstruction error of the entire dose map was used according to

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

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.

Parameter evaluation was performed as follows: First from the 3D x-ray CT images of all 26 patients, 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.

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 slice-wise Greens-function-in-tension-based bidimensional ensemble empirical mode decomposition (GiT-BEEMD) analysis (AL-Baddai *et al* 2016). Details of the algorithm are given in the appendix and a thorough discussion of the impact of different parameters is given in AL-Baddai *et al* (2016). 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 figure 1 are illustrated in figure 4. The idea behind using a GiT-BEEMD and the related IMs is the following: CNNs generate feature maps which are composed of textures on characteristic spatial scales. But this is what an EMD also achieves with its related IMs. Hence it might be profitable to present partial density maps, which only contain characteristic textures instead of complete density maps with all possible textures. This might ease the learning process and help to reduce the variance of the output data arrays.

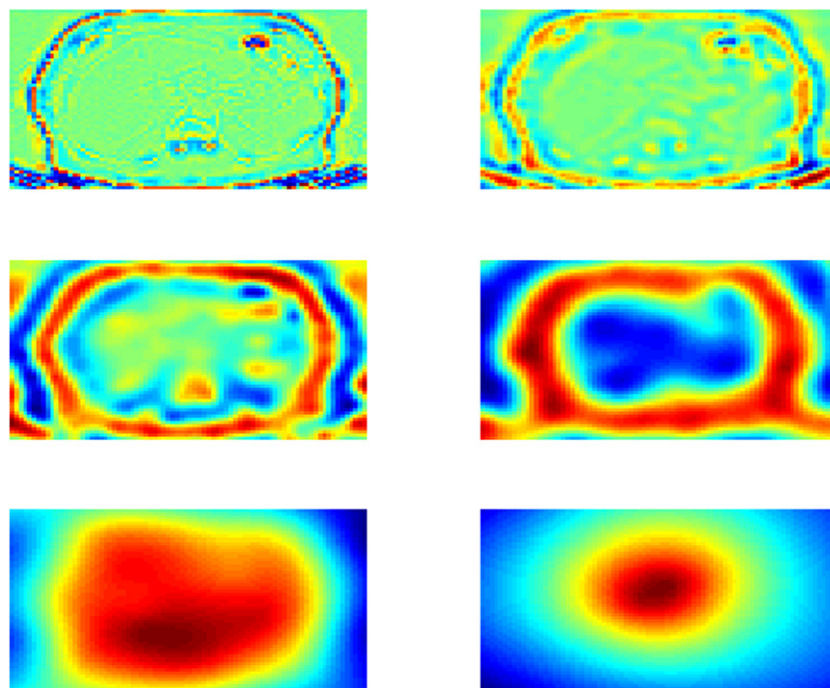


Figure 4. The six intrinsic modes of a slice of the mass density map of a prostate cancer patient.

3. Results

The U-net was trained in two different ways. First, density maps and MIRD-based dose maps were used as inputs to the network and all weights have been trained accordingly. Second, density IMs replaced the density maps, either individually or in specified combinations. The following table 1 provides an overview of the computational load of each method.

3.1. Density maps and MIRD-based dose maps as input

First, one of the input channels provided density maps. During training, the MSE loss function was minimized, and after every epoch, the validation loss was recorded for different network configurations. After training has finished, the test loss was calculated for 350 datasets taken from all 6 test patients and a grand average was computed. The resulting re-scaled losses and their standard deviations are collected in table 2. Note that all input arrays to the hybrid NNET have been normalized by the proper maximal entry in each array. Re-scaling the computed losses has then been achieved by employing an average number of radioactive decays for the estimation of the related dose values.

The testing loss $L = 1.0 \cdot 10^{-2} \text{ Gy}^2 = (1.0 \cdot 10^{-1})^2 \text{ Gy}^2$ 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 losses are marked with a star in the table. The missing statistical significance of most of these values comes because of the small number of patients in the cohort. Still this minimal loss configuration was used for all subsequent calculations.

The following table 3 collects grand averages of the dose values predicted by the different estimation methods discussed. Thereby ‘organwise’ indicates that dose values were first averaged across an organ and then the grand average across the patient cohort has been computed. In case of a ‘voxelwise’ estimation, an average of every voxel dose value across the patient cohort has been determined before dose values were averaged across an organ.

3.2. IMs and MIRD-based dose maps as inputs

Second, 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 ($0.7 \cdot 10^{-2} \text{ Gy}^2$) and validation ($0.9 \cdot 10^{-2} \text{ Gy}^2$) 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.

Table 1. Computational time of the different methods. For the MC simulations a compute cluster with 260 kernels was used. The other methods were calculated on a PC with 8 CPUs.

	MC	NNET	EMD	MIRD
Computational time per patient	ca.6 h	1.3 s	1.9 h	1.1 s

Table 2. Testing losses are given in units of 10^{-2} Gy^2 for different hyperparameter settings. FS denotes filter size and NF gives the number of filters used. The stars mark a significant difference between one 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	1.0 ± 1.2	1.4 ± 2.1
$(9 \times 9 \times 5)$	1.5 ± 2.2	$2.0 \pm 2.7 *$	3.4 ± 2.0

Table 3. Absolute dose values as predicted by the different methods. The values represent averages over the patient cohort. Values in parentheses represent minimal and maximal observed dose values, respectively.

Organwise dose value			
Organ	Monte Carlo (Gy)	NNET/EMD (Gy)	MIRD-based (Gy)
Left kidney	$4.5 \pm 02.3(1.2-08.5)$	$4.3 \pm 02.1(1.2-08.0)$	$5.3 \pm 02.6(1.5-09.7)$
Right kidney	$4.8 \pm 03.2(0.3-13.3)$	$4.6 \pm 03.0(0.3-13.0)$	$5.7 \pm 03.8(0.3-16.3)$
Spleen	$3.1 \pm 03.6(0.1-13.8)$	$2.9 \pm 03.2(0.1-12.3)$	$3.7 \pm 04.1(0.1-15.9)$
Tumor	$12.9 \pm 12.7(0.9-48.4)$	$12.3 \pm 12.2(0.9-47.1)$	$15.7 \pm 15.8(1.2-60.4)$
Voxelwise dose value			
Organ	Monte Carlo (Gy)	NNET/EMD (Gy)	MIRD-based (Gy)
Left kidney	$4.3 \pm 02.9(0.5-032.9)$	$4.1 \pm 02.6(0.5-029.7)$	$5.2 \pm 03.2(0.6-028.0)$
Right kidney	$4.7 \pm 03.6(0.1-036.2)$	$4.5 \pm 03.4(0.1-032.5)$	$5.6 \pm 04.2(0.2-043.4)$
Spleen	$3.0 \pm 04.2(0.1-043.0)$	$2.7 \pm 03.8(0.1-037.2)$	$3.6 \pm 04.7(0.1-038.7)$
Tumor	$14.4 \pm 15.1(0.0-169.5)$	$13.8 \pm 14.5(0.4-159.5)$	$17.6 \pm 18.5(0.3-207.2)$

3.3. Network validation

In order to quantify exactly how well organ doses can be estimated with this method, a leave-one-out cross-validation (LOOCV) was carried out as an alternative to the parameter estimation including training, validation and testing as described above.

The dataset comprised 26 patients. Depending on the target volume-of-interest (VOI), between 27 and 96 adjacent image slices were considered. A leave-one-out cross-validation strategy was employed to follow the generalization ability of the modified U-net. For all patients, the difference of the mean dose within the VOIs predicted either with the modified U-net or the MIRD method, employing a soft tissue kernel, was determined with respect to a full whole-body MC simulation.

The grand averages of the percentage errors of the mean doses, predicted for several target organs, are summarized in table 4 and illustrated with corresponding boxplots in figure 5. Related dose voxel histograms (DVH) are presented in figure 6. Finally per tissue-type average percentage deviations $\Delta D(\%)$ between dosimetry based on full MC simulations and on NNET/EMD are presented in table 5.

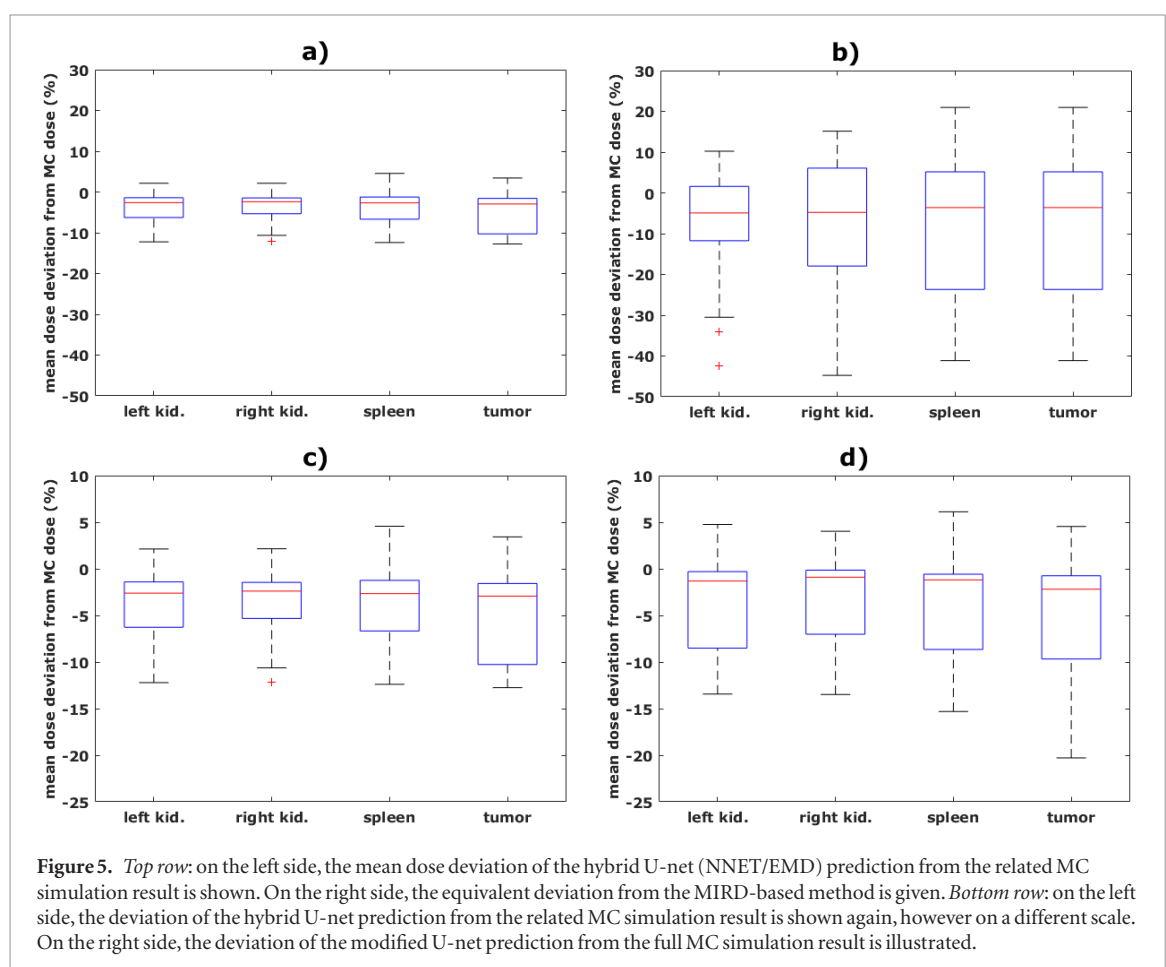
4. Discussion

The current study proposed a hybrid method based on a deep U-Net architecture in combination with an empirical mode decomposition to efficiently estimate radiation dose within an internal radiation therapy. The U-Net possessed two 3D input channels and one 2D output channel. Through the input channels it either received density maps and rough estimates of activity maps, as obtained from a MIRD-based protocol, or the density maps were replaced by one of their intrinsic modes. SPECT/CT data from 26 patients was available for training, validation and testing the proposed model. Model parameters resulted from minimizing a mean squared error loss function, whereby the error measured the deviation of the estimated dose from the corresponding reference dose resulting from individual whole-body Monte Carlo simulations.

EMD is a data-driven method for decomposing spatial density distributions into underlying intrinsic textures with characteristic spatial frequency bands. Usually only few intrinsic modes are extracted, of which the highest spatial frequency mode often mainly contains noise contributions and the lowest spatial frequency

Table 4. Organ-specific percentage errors of mean dose values, averaged over the entire patient cohort.

Organwise dose deviation		
Organ	NNET/EMD—MC (%)	MIRD-based—MC (%)
Left kidney	$-3.7 \pm 4.4 (-12.2-2.1)$	$-7.3 \pm 13.9 (-42.4-10.2)$
Right kidney	$-3.8 \pm 4.3 (-12.2-2.2)$	$-8.5 \pm 16.2 (-44.8-15.1)$
Spleen	$-3.7 \pm 4.7 (-12.4-4.6)$	$-7.3 \pm 16.6 (-41.1-21.0)$
Tumor	$-4.6 \pm 4.8 (-12.8-3.4)$	$-7.3 \pm 16.5 (-40.9-20.8)$
Voxelwise dose deviation		
Organ	NNET/EMD—MC (%)	MIRD-based—MC (%)
Left kidney	$-2.4 \pm 16.3 (-71.0-230.2)$	$24.1 \pm 17.1 (-49.8-658.3)$
Right kidney	$-2.6 \pm 14.9 (-70.9-299.1)$	$23.8 \pm 14.4 (-49.8-658.3)$
Spleen	$-2.7 \pm 15.3 (-84.8-341.9)$	$24.4 \pm 18.3 (-62.1-129.8)$
Tumor	$56.3 \pm 38.9 (-95.8-568.9)$	$87.4 \pm 54.2 (-65.3-756.3)$



modes reflect some general background contribution or any trend in the data. Hence only few modes extract informative textures characteristic for the problem at hand. Presenting only this relevant information instead of all the many irrelevant and distracting details at multiple spatial frequency scales should help to reduce the variance in the data and ease the learning process during training the network.

Because of the small patient cohort, the results presented above can only offer a proof of principle. Still they demonstrate that the network is indeed able to learn an accurate dose estimate. With more data, corresponding estimates could be significantly improved, of course. The reasons for the small number of patients are that neuroendocrine tumors are very rare and a PSMA-Lu therapy is not yet approved in Germany and has only recently been introduced into clinical routine.

The predicted doses, obtained with the modified U-net, are superior to those achieved with the MIRD-based method. In fact, the resulting percentage error is in the range of the statistical error of the MC simulation. Next we compare the LOOCV dose prediction errors obtained either with the modified U-net or the hybrid U-net/

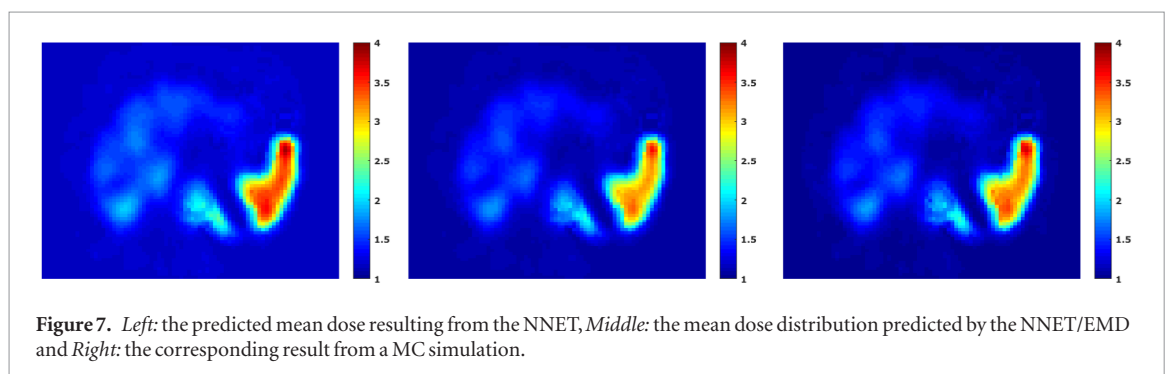
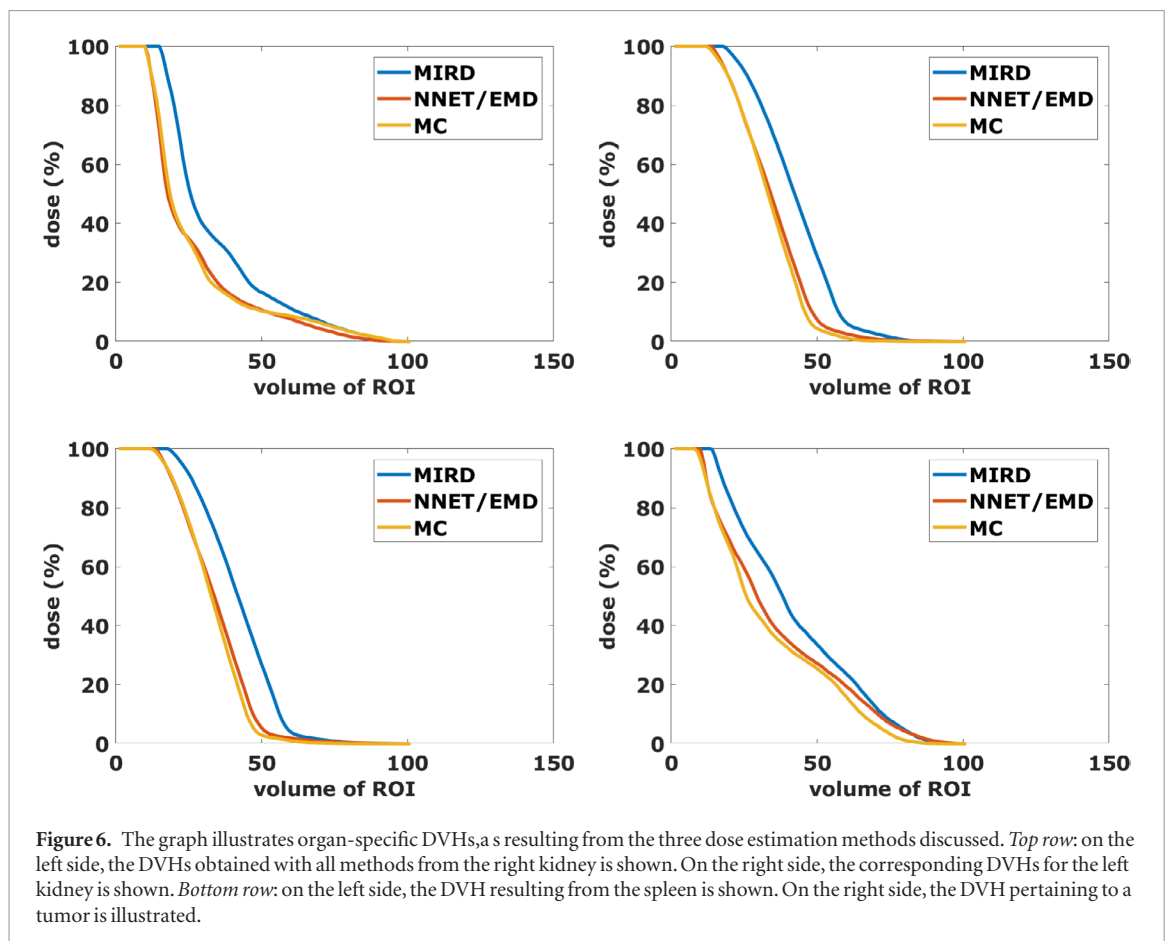


Table 5. Per tissue-type average percentage deviations between dosimetry based on full MC simulations and on NNET/EMD. The numbers indicate relative deviations averaged over the population of 26 patients \pm one standard-deviation.

	ΔD %	Min	Max
Air	0.9 ± 2.3	-1.8	10.3
Lung	2.3 ± 2.0	-0.8	10.0
Adipose (fat)	1.2 ± 2.6	-3.1	7.2
Breast	-0.3 ± 2.8	-5.6	7.6
Soft tissue	-1.7 ± 3.4	-8.1	8.7
Liver	-2.7 ± 4.1	-12.2	9.5
Muscle	-2.4 ± 4.2	-13.6	10.5
Bone compact	-1.0 ± 5.1	-13.0	16.4
Bone cortical	1.7 ± 5.4	-3.6	25.0

EMD relative to the full-body MC dose prediction. The mean predicted dose values are close to each other, but the range of fluctuations around the mean is larger for the modified U-net than for the hybrid U-net/EMD preprocessing. In figure 7, the mean absorbed dose resulting from either the hybrid U-net/EMD or a full MC simulation is illustrated. The dose distributions are nearly identical, even quantitatively.

5. Conclusion

The goal of this study was to improve upon the clinically most often used MIRD-based dose estimations during a radionuclide therapy in nuclear medicine. Our main contribution is a hybrid U-net/EMD architecture, used to predict dose maps from given mass density distributions and activity maps. We compare the predicted doses for several organs with those resulting from the MIRD DVK dose calculation with soft tissue DVK and with those computed with a full-body MC simulation, taken as reference. The results obtained with the new hybrid U-net/EMD are by far superior to the MIRD protocol with soft tissue DVK as it is applied routinely in hospitals. Results obtained with GiT-BEEMD preprocessing indicate an accuracy of the predicted dose similar to the one achieved without preprocessing, but with much less variance. Obviously, the preprocessing removes distracting or redundant information from the input channels. Thus the new technique achieves a precision close to what results from MC dose estimations with little variance in the predicted doses. It improves MIRD-based dose estimation methods considerably and would be easy to apply in daily clinical practice if a model could be trained on a large enough data sample. The latter would need to be rich enough to encompass a large diversity of individual pharmacokinetics and anatomies. In addition transfer learning could be invoked to fine-tune the model to serve a patient—specific radionuclide therapy. In summary, the study represents a proof of concept corroborating that deep learning strategies based on U-Net architectures, and enriched with EMD techniques, can learn to predict absorbed energy dose with voxel spatial resolution and a precision close to MC estimates, but with less computational effort.

Appendix

For the convenience of the reader, a short summary of the GiT-BEEMD is presented.

Empirical mode decomposition (EMD) (Huang *et al* 1998, Zeiler *et al* 2011) tries to locally extract from any multi-variate signal simple oscillatory components, called intrinsic modes (IM). In case of two-dimensional intensity distributions such as images, the data sets need to be centered beforehand. The resulting spatial oscillations can be seen as characteristic textures of the given intensity distribution. IMs are obtained from the signal through a process called sifting, which results in pure oscillations $x(t) = a(t) \exp(i\phi(t)) \in \mathbb{C}$ with zero-mean but spatially and/or temporally varying amplitude and frequency. Furthermore, IMs are ordered according to their (spatial)-frequency content, where $k = d\phi(r, t)/dr$ and $\omega(t) = d\phi(r, t)/dt$. Contrary to wavelet analysis, EMD is a data driven algorithm that decomposes the signal without prior knowledge.

The decomposition of a 2D data array starts by vectorizing it into a row vector $\mathbf{x}_m \in \mathbb{R}^N$. All data arrays are then collected into a data matrix $\mathbf{X}^{M \times N}$. First, EMD 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 data arrays and N gives the dimension of the data vectors, i.e. the number of elements of every input data array. The EMD decomposition of the n th column of \mathbf{X} and the subsequent decomposition of the m th row of the resulting component matrices yields

$$X_{*,n} = \sum_{j=1}^J C_{*,n}^{(j)} \quad \text{and} \quad C_{m,*}^{(j)} = \sum_{k=1}^K H_{m,*}^{(j,k)}. \quad (\text{A.1})$$

Applying the *comparable minimal scale principle* (CMSP) (Wu *et al* 2009), i.e. summing up all component matrices $\mathbf{H}^{(j,k)}$, which represent a similar spatial scale, generates intrinsic modes:

$$\mathbf{s}^{(k')} = \sum_{k=1}^K \mathbf{H}^{(k,k')} + \sum_{j=k+1}^J \mathbf{H}^{(j,k')}. \quad (\text{A.2})$$

Finally, we thus obtain an expansion of the data matrix into its underlying intrinsic modes

$$\mathbf{X} = \sum_{k'=1}^K \mathbf{s}^{(k')}. \quad (\text{A.3})$$

A noise—assisted version is obtained by adding and subtracting white noise ϵ with zero mean and variance $\sigma = 0.04$ to the data. Averaging the related noisy IMs helps in avoiding mode mixing, favors denoising the data and results in the final intrinsic modes.

The extracted IMs are highly dependent on the sifting process. The latter is carried out in this study with the help of Green's functions under tension (AL-Baddai *et al* 2016). Local extrema of the 2D intensity distribu-

tion are located employing an eight-connected neighbors approach yielding corresponding maxima and minima maps. Extrema are obtained whenever $D(m, n) = x_{mm}x_{nn} - (x_{mn}^2) > 0$, whereby the signs of x_{mm}, x_{nn} determine whether the point is a maximum or a minimum. These extrema are interpolated smoothly with Green's functions $\phi(\mathbf{x}_u, \mathbf{x}_v)$ to obtain upper and lower envelope surfaces. Points at these envelope surfaces can be expressed as a weighted sum of Green's functions according to

$$s(\mathbf{x}_u) = \sum_{v=1}^V w_v \cdot \phi(\mathbf{x}_u, \mathbf{x}_v). \quad (\text{A.4})$$

Green's functions obey the following equation, evaluated at $V = M \times N$ data constraints

$$\sum_{v=1}^V w_v [D\Delta_{op}^2 - T\Delta_{op}] \phi(\mathbf{x}_u, \mathbf{x}_v) = \sum_{v=1}^V w_v \delta(\mathbf{x}_u - \mathbf{x}_v) \quad (\text{A.5})$$

yielding the following Green's function

$$\phi(\mathbf{x}_u, \mathbf{x}_v) = \log(p \cdot |\mathbf{x}_u, \mathbf{x}_v|) + K_0(p \cdot |\mathbf{x}_u, \mathbf{x}_v|) \quad (\text{A.6})$$

whereby $K_0(\dots)$ denotes a modified Bessel-function of second kind and order zero, D denotes the flexural surface rigidity, T the surface tension and $p = T/D$.

The sifting process thus follows two steps:

Step 1 Weights \mathbf{w} are estimated from

$$\mathbf{G}\mathbf{w} = \mathbf{s} \quad (\text{A.7})$$

where \mathbf{G} denotes the matrix of Green's functions.

Step 2 The envelope surface is interpolated at unknown points $s(\mathbf{x}_u)$ according to

$$s(\mathbf{x}_u) = \mathbf{w}^T \Phi \quad (\text{A.8})$$

where $\Phi = [\Phi(\mathbf{x}_u, \mathbf{x}_1), \dots, \Phi(\mathbf{x}_u, \mathbf{x}_V)]^T$.

ORCID iDs

Th I Götz  <https://orcid.org/0000-0001-8751-3404>

E W Lang  <https://orcid.org/0000-0001-7440-0224>

References

- AL-Baddai S M, Al-Subar K, Tomé A M, Solé-Casals J and Lang E W 2016 A Green's function-based bi-dimensional empirical mode decomposition *Inf. Sci.* **348** 305–21
- Arce P *et al* 2014 Gamos: a framework to do GEANT4 simulations in different physics fields with an user-friendly interface *Nucl. Instrum. Methods Phys. Res. A* **735** 304–13
- Bolch W E *et al* 1999 MIRD pamphlet no. 17: the dosimetry of nonuniform activity distributions—radionuclide s-values at the voxel level *J. Nucl. Med.* **40** 11S–36S
- Botta F *et al* 2011 Calculation of electron and isotopes dose point kernels with FLUKA Monte Carlo code for dosimetry in nuclear medicine therapy *Med. Phys.* **38** 3944–54
- Cal-Gonzalez J, Rausch I, Sundar L K S, Lassen M L, Muzik O, Moser E, Papp L and Beyer T 2018 Hybrid imaging: instrumentation and data processing *Front. Phys.* **6** 1–30
- Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J and Wang G 2017 Low-dose CT denoising with convolutional neural network *IEEE 14th Int. Symp. on Biomedical Imaging* pp 143–6
- Götz T I 2019b Technical report: time-activity-curve integration in Lu-177 therapies in nuclear medicine (arXiv:1907.06617)
- Götz T I, Lang E W, Prante O, Cordes M, Kuwert T, Maier A, Ritt P and Schmidkonz C 2019 Estimation of 177Lu-PSMA-617 tumor uptake based on three-dimensional Monte Carlo based voxel-wise tumor dosimetry in patients with metastasized castration resistant prostate cancer *Ann. Nucl. Med.* under revision
- Götz T I, Lang E W, Schmidkonz C, Maier A, Kuwert T and Ritt P 2019 Particle filter de-noising of voxel-specific time-activity-curves in personalized Lu-177 therapy *Z. Med. Phys.* (accepted) (<https://doi.org/10.1016/j.zemedi.2019.10.005>)
- Götz T I, Schmidkonz C, Lang E W, Maier A, Kuwert T and Ritt P 2019a A comparison of methods for adapting 177Lu dose-voxel-kernels to tissue inhomogeneities *Phys. Med. Biol.* **64** 245011
- Götz T I, Schmidkonz C, Lang E W, Maier A, Kuwert T and Ritt P 2019b Factors affecting accuracy of S-values and determination of time-integrated activity in clinical Lu-177 dosimetry *Ann. Nucl. Med.* **33** 521–31
- Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C C and Liu H H 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis *Proc. R. Soc. A* **454** 903–95
- Jackson P, Hardcastle N, Dawe N, Kron T, Hofman M S and Hicks R J 2018 Deep learning renal segmentation for fully automated radiation dose estimation in unsealed source therapy *Front. Oncol.* **8** 215–22

- Kalantzis G, Vasquez-Quino L A, Zalman T, Pratz G and Lei Y 2011 Toward IMRT 2D dose modeling using artificial neural networks: a feasibility study *Med. Phys.* **38** 5807–17
- Kearney V, Chan J, Descovich M, Yom S S and Solberg T D 2018a A multi-task CNN model for autosegmentation of prostate patients *Radiat. Oncol. Biol. Phys.* **102** S214
- Kearney V, Chan J, Haaf S, Descovich M and Solberg T D 2018b DoseNet: a volumetric dose prediction algorithm using 3D fully-convolutional neural networks *Phys. Med. Biol.* **63** 235022
- Kearney V P, Haaf S, Sudhyadhom A, Valdes G and Solberg T D 2018c An unsupervised convolutional neural network-based algorithm for deformable image registration *Phys. Med. Biol.* **63** 185017
- Lee M S, Hwang D, Kim J H and Lee J S 2019 Deep-dose: a voxel dose estimation method using deep convolutional neural network for personalized internal dosimetry *Nature* **9** 10308
- Lou B, Doken S, Zhuang T, Wingerter D, Gidwani M, Mistry N, Ladic L, Kamen A and Abazeed M E 2019 An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction *Lancet Digit. Health* **1** e136–47
- Nguyen D, Long T, Jia X, Lu W, Gu X, Iqbal Z and Jiang S 2017 Dose prediction with U-net: a feasibility study for predicting dose distributions from contours using deep learning on prostate IMRT patients (arXiv:1709.09233)
- Pacilio M et al 2015 Differences in 3D dose distributions due to calculation method of voxel s-values and the influence of image blurring in SPECT *Phys. Med. Biol.* **60** 1945
- Papadimitroulas P, Loudos G, Nikiforidis G and Kagadis G 2012 A dose point kernel database using gate Monte Carlo simulation toolkit for nuclear medicine applications: comparison with other Monte Carlo codes *Med. Phys.* **39** 5238–47
- Park J, Hwang D, Kim K Y, Kang S K, Kim Y K and Lee J S 2018 Computed tomography super-resolution using deep convolutional neural network *Phys. Med. Biol.* **63** 145011
- Pérez P, Botta F, Pedroli G and Valente M 2011 *Dosimetry for Beta-Emitter Radionuclides by Means of Monte Carlo Simulations* (Rijeka: IntechOpenY) (<https://doi.org/10.5772/25287>)
- Petoussi-Henss N, Bolch W, Zankl M, Sgouros G and Wessels B 2007 Patient-specific scaling of reference s-values for cross-organ radionuclide s-values: what is appropriate? *Radiat. Prot. Dosim.* **127** 192–6
- Ronneberger O, Fischer P and Brox T 2015 *U-Net: Convolutional Networks for Biomedical Image Segmentation* ed N Navab et al (New York: Springer) pp 234–41
- Sanders J, Kuwert T, Hornegger J and Ritt P 2015 Quantitative spect/ct imaging of ¹⁷⁷Lu with *in vivo* validation in patients undergoing peptide receptor radionuclide therapy *Mol. Imaging Biol.* **17** 585–93
- Schneider W, Bortfeld T and Schlegel W 2000 Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions *Phys. Med. Biol.* **45** 459–78
- Wu Z, Huang N E and Chen X 2009 The multi-dimensional ensemble empirical mode decomposition method *Adv. Adapt. Data Anal.* **1** 339–72
- Zeiler A, Faltermeier R, Tomé A M, Puntonet C G, Brawanski A and Lang E W 2011 Sliding empirical mode decomposition for on-line analysis of biomedical time series *Neural Process. Lett.* **37** 299–306