

M2 Internship (with PhD prospects) Optimal Transport for Dose Estimation in Cancer Radiotherapy

CentraleSupélec Rennes – IETR

Pursuing a PhD is possible following the internship. Applications for the PhD position will also be considered. Feel free to reach out for any information regarding the thesis.

General Context

Radiotherapy is one of the main treatments for cancer. It involves delivering a precise dose of radiation to destroy tumor cells while preserving healthy tissue. The quantitative evaluation of this dose, known as *dosimetry*, relies on dose maps that indicate, for each point in the body, the amount of absorbed energy. These maps are typically estimated using *Monte Carlo* (MC) simulations, known for their accuracy but computationally expensive due to their iterative and stochastic estimation of probability density functions.

To accelerate these computations, several techniques have been developed: variance reduction methods [1, 2], GPU acceleration [3, 4], and the use of *Deep Learning* (DL) to partially replace MC simulation. Recent studies [5–7] have demonstrated the feasibility of convolutional networks capable of rapidly estimating dose distributions, even in real time [8–10].

However, despite these advances, DL models remain difficult to deploy in clinical settings: their training phase is complex, they poorly integrate physical knowledge, and they still lack robustness and precision for reliable use in medical practice.

Improving dose estimation accuracy and reducing prediction uncertainty under all clinical conditions remain major challenges in medical physics. To address this, we propose leveraging *Optimal Transport* (OT) in various aspects of dose estimation.

OT provides a rigorous mathematical framework for establishing correspondences and measuring distances between probability distributions by quantifying their spatial discrepancies. This leads to the definition of distribution distances, such as the Wasserstein distance [11]. In medical dosimetry, OT can be used to :

- detect subtle biases and more finely measure similarity between dose maps, offering more reliable evaluation than classical metrics;
- guide model transfer to achieve robust predictions on new clinical datasets, even when training data is limited.

This approach aims to enhance the accuracy and robustness of dose maps while paving the way for better clinical interpretation.

Internship Objectives

This internship is part of the ExOTics project (ANR JCJC), funded by the French National Research Agency (ANR). It represents the first step in a research effort to apply OT to dose evaluation in radiotherapy. The internship will explore the practical potential of OT on simulated dose maps before integrating it into DL models.

Phase 1 - Exploration and Implementation of Existing Metrics

- Conduct a literature review of metrics used to compare dose maps (MAE, MSE, gamma index, SSIM, etc.),

- Understand the theoretical foundations of optimal transport and the Wasserstein distance,
- Propose and implement in Python a new metric based on the Wasserstein distance for comparing simulated dose maps,
- Study the impact of statistical variability (low-sampling regions) on the performance of these metrics,
- Provide a critical analysis of the limitations of classical approaches and the benefits of OT.

Phase 2 – Advanced Exploration and Research Directions Direct computation of the Wasserstein distance from Monte Carlo samples remains time- and memory-intensive. This phase will explore more efficient approaches for large-scale use:

- Investigate approximations based on statistical moments (mean, variance, etc.) to reduce computational cost,
- Evaluate moment-based Wasserstein distance approximations inspired by Mula & Nouy (2024) [12],
- Adapt these approaches to the discrete framework of dose maps and propose a simplified Python implementation,
- Test performance on simulated maps with varying noise levels.

Student Profile

Master’s student or engineering school student with a research focus. Candidates should be team-oriented, autonomous, and proactive. Desired technical skills include a generalist background with coursework in machine learning and/or applied mathematics, strong proficiency in Python, solid scientific English, and proven scientific writing abilities.

Practical Information

- **Duration** : 4 to 6 months (start between February–April depending on availability)
- **Location** : CentraleSupélec Rennes – Beaulieu campus
- **Compensation** : ~ 600 euros/month
- **Supervisors** : Hong-Phuong Dang & Clément Elvira
- **Contact**: Send to hong-phuong.dang@centralesupelec.fr : CV + cover letter + transcripts from the last three years

References

- [1] A. Behloul, D. Visvikis, and J. Bert. Improved woodcock tracking on monte carlo simulations for medical applications. *Physics in medicine and biology*, 63, 2018.
- [2] S. García-Pareja, A. Lallena, and F. Salvat. Variance-reduction methods for monte carlo simulation of radiation transport. *Frontiers in Physics*, 9, 2021.
- [3] J. Bert, H. Perez-Ponce, Z. El Bitar, S. Jan, Y. Boursier, D. Vintache, A. Bonissent, C. Morel, D. Brasse, and D. Visvikis. Geant4-based monte carlo simulations on gpu for medical applications. *Physics in medicine and biology*, 58, 2013.

- [4] D. Maneval, B. Ozell, and P. Després. pgpumcd: an efficient gpu-based monte carlo code for accurate proton dose calculations. *Physics in medicine and biology*, 64, 2019.
- [5] Z. Liu, J. Fan, M. Li, Y. Han, Z. Hu, P. Huang, Y. Tian, J. Miao, and J. Dai. A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy. *Medical physics*, 46, 2019.
- [6] C. Kontaxis, G. Bol, J. Lagendijk, and B. Raaymakers. Deepdose: Towards a fast dose calculation engine for radiation therapy using deep learning. *Physics in medicine and biology*, 65, 2020.
- [7] M. Villa, B. Nasr, D. Benoit, N. Padoy, D. Visvikis, and J. Bert. Fast dose calculation in x-ray guided interventions by using deep learning. *Physics in medicine and biology*, 68, 2023.
- [8] A. Touil, H.-P. Dang, D. Benoit, D. Visvikis, and J. Bert. Toward a generic approach for dose distribution recovery by using deep learning and gpu-based geant4 simulations. In *The IV Geant4 International User Conference at the physics-medicine-biology frontier*, 2022.
- [9] H.-P. Dang, J. Bert, and D. Visvikis. Improving statistical quality and computational time with adapting deep learning for monte carlo dose distribution. 2022.
- [10] D. Sarrut, A. Etxebeste, N. Krah, and J. Létang. Modeling complex particles phase space with gan for monte carlo spect simulations: a proof of concept. *Physics in medicine and biology*, 66, 2021.
- [11] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11, 2019.
- [12] O. Mula and A. Nouy. Moment-sos methods for optimal transport problems. *Numerische Mathematik*, 2024.

Stage M2 (avec perspective de thèse)

Transport optimal pour l'estimation de la dose délivrée en radiothérapie dans le traitement du cancer

CentraleSupélec Rennes – IETR

Une poursuite en thèse est envisageable à l'issue du stage. Les candidatures pour l'offre de thèse seront également étudiées. N'hésitez pas à prendre contact pour toute information concernant la thèse.

Contexte général

La *radiothérapie* est l'un des traitements majeurs du cancer. Elle consiste à délivrer une dose précise de rayonnement pour détruire les cellules tumorales tout en préservant les tissus sains. L'évaluation quantitative de cette dose, appelée *dosimétrie*, repose sur des cartes de dose indiquant, pour chaque point du corps, la quantité d'énergie absorbée. Ces cartes sont généralement estimées par des simulations Monte-Carlo (MC), reconnues pour leur précision mais très coûteuses en calcul, car elles reposent sur des estimations itératives et stochastiques de fonctions de densité de probabilité.

Pour accélérer ces calculs, plusieurs techniques ont été développées : méthodes de réduction de variance [1, 2], exploitation des GPU [3, 4], ou encore recours à l'*apprentissage profond* (*Deep Learning*, DL) pour remplacer partiellement la simulation MC. Des travaux récents [5–7] ont démontré la faisabilité de réseaux convolutifs capables d'estimer rapidement des distributions de dose, voire en temps réel [8–10].

Cependant, malgré ces progrès, les modèles DL restent difficiles à exploiter en clinique : leur phase d'apprentissage est complexe, ils intègrent mal les connaissances physiques et manquent encore de robustesse et de précision pour une utilisation fiable en pratique médicale.

Améliorer la précision de l'estimation de la dose et réduire l'incertitude des prédictions dans toutes les conditions cliniques restent des défis majeurs en physique médicale. Pour y répondre, nous proposons d'exploiter le *transport optimal* (*Optimal Transport*, OT) dans différents volets de l'estimation des doses.

L'OT fournit un cadre mathématique rigoureux pour établir des correspondances et mesurer des distances entre des distributions de probabilité en quantifiant leurs écarts à l'aide de mesures spatiales. Cette approche conduit à définir des distances entre distributions, telles que la distance de Wasserstein [11]. Dans le contexte de la dosimétrie médicale, l'OT peut servir à :

- détecter des biais subtils et mesurer plus finement la similarité entre cartes de dose, pour une évaluation plus fiable que les métriques classiques ;
- guider le transfert de modèles afin d'obtenir des prédictions robustes sur de nouveaux jeux de données cliniques, même lorsque les données d'entraînement sont limitées.

Cette approche vise à renforcer l'exactitude et la robustesse des cartes de dose, tout en ouvrant la voie à une meilleure interprétation clinique.

Objectifs du stage

Ce stage s'inscrit dans le cadre du projet ExOTics (ANR JCJC) financé par l'Agence nationale de la recherche (ANR). Dans ce cadre, le stage proposé constitue la première étape d'un travail de recherche visant à exploiter l'OT pour l'évaluation de dose en radiothérapie. Il permettra d'explorer concrètement le potentiel de l'OT sur des cartes de dose simulées avant son intégration dans des modèles de DL.

Première phase – Exploration et mise en œuvre des métriques existantes.

- Réaliser un état de l'art des métriques utilisées pour comparer les cartes de dose (MAE, MSE, gamma index, SSIM, etc.) ;

- Assimiler les fondements théoriques du transport optimal et de la distance de Wasserstein ;
- Proposer et implémenter en Python une nouvelle métrique basée sur la distance de Wasserstein pour la comparaison de cartes de dose simulées ;
- Étudier l’impact de la variabilité statistique (zones à faible échantillonnage) sur les performances de ces différentes métriques ;
- Proposer une analyse critique des limites des approches classiques et des bénéfices de l’OT.

Deuxième phase – Approfondissement et pistes de recherche. Le calcul direct de la distance de Wasserstein à partir des échantillons Monte-Carlo reste coûteux en temps et en mémoire. La seconde phase visera donc à explorer des approches plus efficaces pour une utilisation à grande échelle :

- Étudier des approximations basées sur les moments statistiques (moyenne, variance, etc.) pour réduire le coût de calcul ;
- Évaluer les méthodes d’approximation de la distance de Wasserstein à partir de moments inspirées de Mula & Nouy (2024) [12] ;
- Adapter ces approches au cadre discret des cartes de dose et en proposer une implémentation simplifiée (Python) ;
- Tester les performances sur des cartes simulées avec différents niveaux de bruit.

Profil souhaité

Étudiant(e) en école d’ingénieur et/ou en Master Recherche. Les candidat(e)s doivent avoir une aptitude au travail en équipe, être autonomes et animé(e)s d’un esprit d’initiative. Les compétences techniques souhaitées sont une formation généraliste avec un cursus en apprentissage automatique et/ou en mathématiques appliquées, une bonne maîtrise du langage Python, ainsi qu’un bon niveau en anglais scientifique et des compétences avérées en rédaction scientifique.

Informations pratiques

- **Durée** : 4 à 6 mois (démarrer entre février–avril selon le souhait)
- **Lieu** : CentraleSupélec Rennes – campus de Beaulieu
- **Gratification** : selon réglementation (~ 600 euros/mois)
- **Encadrement** : Hong-Phuong Dang & Clément Elvira
- **Contact** : Envoyer CV+lettre de motivation + bulletins de notes des trois dernières années au contact : hong-phuong.dang@centralesupelec.fr

Références

- [1] A. Behloul, D. Visvikis, and J. Bert. Improved woodcock tracking on monte carlo simulations for medical applications. *Physics in medicine and biology*, 63, 2018.
- [2] S. García-Pareja, A. Lallena, and F. Salvat. Variance-reduction methods for monte carlo simulation of radiation transport. *Frontiers in Physics*, 9, 2021.
- [3] J. Bert, H. Perez-Ponce, Z. El Bitar, S. Jan, Y. Boursier, D. Vintache, A. Bonissent, C. Morel, D. Brasse, and D. Visvikis. Geant4-based monte carlo simulations on gpu for medical applications. *Physics in medicine and biology*, 58, 2013.
- [4] D. Maneval, B. Ozell, and P. Després. pgpumcd : an efficient gpu-based monte carlo code for accurate proton dose calculations. *Physics in medicine and biology*, 64, 2019.
- [5] Z. Liu, J. Fan, M. Li, Y. Han, Z. Hu, P. Huang, Y. Tian, J. Miao, and J. Dai. A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy. *Medical physics*, 46, 2019.

- [6] C. Kontaxis, G. Bol, J. Legendijk, and B. Raaymakers. Deepdose : Towards a fast dose calculation engine for radiation therapy using deep learning. *Physics in medicine and biology*, 65, 2020.
- [7] M. Villa, B. Nasr, D. Benoit, N. Padoy, D. Visvikis, and J. Bert. Fast dose calculation in x-ray guided interventions by using deep learning. *Physics in medicine and biology*, 68, 2023.
- [8] A. Touil, H.-P. Dang, D. Benoit, D. Visvikis, and J. Bert. Toward a generic approach for dose distribution recovery by using deep learning and gpu-based geant4 simulations. In *The IV Geant4 International User Conference at the physics-medicine-biology frontier*, 2022.
- [9] H.-P. Dang, J. Bert, and D. Visvikis. Improving statistical quality and computational time with adapting deep learning for monte carlo dose distribution. 2022.
- [10] D. Sarrut, A. Etxebeste, N. Krah, and J. Létang. Modeling complex particles phase space with gan for monte carlo spect simulations : a proof of concept. *Physics in medicine and biology*, 66, 2021.
- [11] G. Peyré and M. Cuturi. Computational optimal transport : With applications to data science. *Foundations and Trends in Machine Learning*, 11, 2019.
- [12] O. Mula and A. Nouy. Moment-sos methods for optimal transport problems. *Numerische Mathematik*, 2024.