Physics and Imaging in Radiation Oncology 37 (2026) 100916

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Original Research Article

Inverse consistency error for validating deformable image registration: an explorative study on computational phantoms

Gianfranco Loi [a], Marco Fusella [b,*], Stefania Zara [c], Marica Vagni [d], Nicola Michielli [e], Orlando Zaccaria [c], Lorenzo Placidi [f], Pierfrancesco Franco [a,g], Filippo Molinari [e], Christian Fiandra [h]

[a] Azienda Ospedaliero-Universitaria Maggiore della Carità, Novara, Italy
[b] Abano Terme Hospital, Department of Radiation Oncology, Abano Terme, Padua, Italy
[c] Tecnologie Avanzate Srl, Turin, Italy
[d] Università Cattolica del Sacro Cuore, Campus di Roma, Rome, Italy
[e] PoliToBIOMed Lab, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy
[f] Diagnostic Imaging and Oncological Radiotherapy Department Fondazione Policlinico Universitario A. Gemelli, Rome, Italy
[g] Università degli Studi del Piemonte Orientale "Amedeo Avogadro", Novara, Italy
[h] Azienda Ospedaliero Universitaria (AOU) Città della Salute e della Scienza, Presidio Molinette, Turin, Italy

ARTICLE INFO

Keywords:
Deformable image registration
Inverse consistency error
Registration uncertainty
Voxel wise validation
Quality assurance

ABSTRACT

Background and Purpose: Validation of deformable image registration (DIR) remains predominantly contourbased; this study evaluated inverse consistency error (ICE) as an automated voxelwise metric for DIR accuracy.

Materials and Methods: Synthetic ground-truth DVFs were generated using geometric and head-and-neck (HN) digital phantoms undergoing controlled global and local deformations. DIR was performed with the ANACONDA algorithm in RayStation. ICE maps derived from clinical DVFs were compared with ground-truth registration error (GTRE), target registration error (TRE) from 20 anatomical landmarks, and mean distance to agreement (MDA) for 22 propagated ROIs.

Results: Ground-truth DVFs showed negligible ICE values, confirming mathematical invertibility. In HN phantoms, median ICE and GTRE were 0.8 ± 0.2 mm and 1.6 ± 0.4 mm, respectively. ICE correlated strongly with GTRE ($R = 0.85$, $p < 0.001$) and moderately with TRE ($R = 0.68$, $p < 0.001$). No significant correlation was found with contourbased MDA (2.47 ± 0.18 mm). Voxel-wise analysis showed that ICE captured spatial patterns of uncertainty consistent with regions of higher GTRE, while underestimating error for global homogeneous deformations >15 mm due to DIR regularisation. Across all datasets, ICE correctly identified high-uncertainty subregions that were not detected by contour-based metrics.

Conclusions: ICE enables automated voxel-wise quantification of DIR uncertainty directly from clinical DVFs. It complements traditional contour-based metrics and may support patient-specific QA and more reliable dose mapping in adaptive and re-irradiation radiotherapy workflows.

1. Introduction

Deformable Image Registration (DIR) is increasingly used in adaptive and re-irradiation workflows. Although commercial treatment planning systems incorporate DIR to account for anatomical changes, they generally operate as "black-box" processes and do not provide voxel-wise uncertainty quantification or dedicated tools for patient-specific quality assurance. This lack of transparency reinforces the need for independent DIR validation strategies. Classical evaluation metrics—such as target registration error (TRE), Dice similarity coefficient (DSC), and mean distance to agreement (MDA)—are widely used to assess contour propagation [1–5]. However, they are labor-intensive, operator-dependent, and not suited for voxel-level validation required for dose mapping and accumulation [6,7]. Moreover, these metrics may fail to detect nonphysical or inconsistent deformations that preserve contours but distort the underlying deformation vector fields (DVFs)

* Corresponding author at: Department of Radiation Oncology, Abano Terme Hospital, Piazza Cristoforo Colombo 1, 35031 Abano Terme, PD, Italy.
E-mail address: mfusella@casacura.it (M. Fusella).

https://doi.org/10.1016/j.phro.2026.100916
Received 16 June 2025; Received in revised form 30 January 2026; Accepted 30 January 2026
Available online 7 February 2026

[8,9]. Jacobian- or strain-based methods can highlight local irregularities but are rarely used as external QA tools and lack intuitive interpretability for routine clinical use [10–13].

Inverse consistency error (ICE) quantifies the deviation between forward and backward DVFs. Although not a sufficient condition for physical plausibility [14], ICE has two appealing properties: it directly exploits the DVFs generated by clinical registration workflows, and it enables automated voxel-wise estimation of local uncertainty without manual input. Prior work has investigated ICE in limited or idealized settings, but its clinical feasibility and predictive value for patient-specific QA remain insufficiently characterized [12,15,16].

The aim of this study is to evaluate the potential of ICE as a validation metric for DIR in a patient-specific setting. Although the use of ICE as an evaluation metric for DIR is not new, this work focuses on its application on clinical DVFs. Specifically, we propose using ICE maps derived from clinical DVFs to estimate geometric uncertainties at voxel level. This approach was tested across a range of controlled deformation scenarios using both geometric and digital head and neck phantoms, with ground-truth DVFs provided by a dedicated deformation simulation platform. ICE was then compared to Ground Truth Registration Errors (GTRE) and alternative metrics (TRE, MDA) to assess its robustness and predictive value. To our knowledge, this is one of the first studies to examine the feasibility of implementing ICE as an independent, automated Quality Assurance (QA) tool for assessing voxel-level DIR accuracy within a clinical workflow.

2. Materials and methods

2.1. Image deformation procedure: ground-truth generation

To simulate realistic but controlled anatomical variations, we used the ImSimQA software package (Oncology Systems Limited, Shrewsbury, UK) to generate synthetic ground-truth deformations and their associated deformation vector fields (DVFs) [14–17]. This platform enables both global and local 3D image deformations using two different interpolation schemes: thin-plate splines (ThPS) [18] and compactly supported radial basis functions (CSRBF) [19], which allow the creation of smooth, invertible transformations [20,21].

Geometric phantoms were created to simulate the controlled transformation of a sphere into an ellipsoid within a cubic volume (512 × 512 × 44 voxels, 0.25 × 0.25 × 3 mm^3). Using the ThPS-based algorithm, six increasing levels of deformation (from 3 mm to 21 mm) were applied along one axis of the sphere to generate global DVFs (Supplementary Materials Figs. E1–E4). For each deformation, both the forward and inverse DVFs were computed and exported.

These phantoms served as an initial testbed to verify DVF invertibility and to calibrate the validation pipeline, including harmonization of sampling grids and displacement units. ICE maps were calculated to confirm that the synthetic transformations satisfied inverse consistency and could be reliably used as ground truth.

To evaluate the method in a clinically relevant setting, we generated synthetic anatomical phantoms inspired by off-line head and neck (HN) adaptive radiotherapy scenarios [6,14]. A planning CT scan of a Head and Neck (HN) patient (512 × 512 × 85 voxels; voxel size 1.27 × 1.27 × 3 mm^3), along with manually segmented organs at risk (OARs), was used as the base image. Three deformed versions of this phantom detailed in Table 1 were created to simulate progressive shrinkage of the parotid glands and tumor volume, using the following deformation strategies: first Global deformation via ThPS (HN_G), then Local deformation using CSRBF interpolation (HN_L) and Local deformation with stiffness constraints to preserve bone surfaces (HN_R) and simulate clinical reliable volume changes during the RT courses (Fig. 1).

For each anatomical phantom, both forward and inverse DVFs were generated and exported. ICE maps were computed to assess mathematical consistency of these synthetic ground-truth transformations.

2.2. The DIR algorithm

Deformable image registration was performed using the ANACONDA algorithm, as implemented in the RayStation treatment planning system (v9.A, RaySearch Laboratories AB, Stockholm, Sweden) [22]. ANACONDA is a hybrid algorithm that incorporates both intensity-based similarity and anatomical feature alignment, using multiple regularization terms to ensure smooth and locally invertible deformations.

In the registration workflow, regularization ROIs were used to guide the alignment: for geometric phantoms, the External and Sphere/Ellipsoid ROIs were defined (Supplementary Fig. E4); for anatomical phantoms, according to the clinical protocol implemented at our respective institutions, the External, Mandible, and Spinal Canal ROIs were employed. All registrations were performed using a 2.5 mm isotropic deformation grid.

2.3. ICE–GTRE comparisons and phantom rationale

ICE was defined as the voxel-wise distance between the composition of the forward and backward DVFs and the identity transformation [12]. DVF composition was computed according to the method described by Bender et al. [15], as detailed in Supplementary Materials, Section A1 and A2. GTRE maps were obtained by composing the forward ground-truth DVF with the backward clinical DVF and comparing the result to the identity transformation (Supplementary Section A3). ICE and GTRE maps were obtained by composing the forward ground-truth DVF with the backward clinical DVF and comparing the result to the identity transformation (Supplementary Section A4).

1. Maximum value and 1D spatial consistency (Fig. 2, HN_G). Maximum ICE and GTRE values and 1D profiles were compared to assess whether ICE captures global error magnitude and large-scale spatial trends. The HN_G phantom, based on a smooth global deformation, represents an idealized scenario suited to testing ICE sensitivity under mathematically favorable conditions.
2. 2D spatial consistency (Fig. 3, HN_L). Voxel-wise 2D ICE and GTRE maps were compared to evaluate spatial correspondence in the presence of localized deformation. The HN_L phantom introduces spatial discontinuities through localized shrinkage, stressing DIR performance and testing ICE's ability to localize registration errors.
3. Maximum error scaling (Fig. 4, geometric phantoms). Maximum ICE and GTRE were compared across geometric phantoms to examine

Table 1
Percentage of GTV and Parotid shrinkage for each deformation strategy and applied deformation.

Deformation strategy	% shrinkage						Deformation strategy (explanatory note)
	DVF1		DVF2		DVF3		
	GTV	Parotids	GTV	Parotids	GTV	Parotids	
HN_G	5%	8–9%	14%	13–15%	33%	17–19%	Global smooth deformation of the entire image domain using ThPs, independent of tissue type; used to test ICE sensitivity under ideal conditions.
HN_L	11%	7–8%	20%	13–14%	35%	18–20%	Localized volume shrinkage in otherwise stable anatomies, introducing spatial discontinuities.
HN_R	7%	6–7%	17%	11–12%	26%	16–17%	Regional deformation with intermediate characteristics between global and localized strategies.

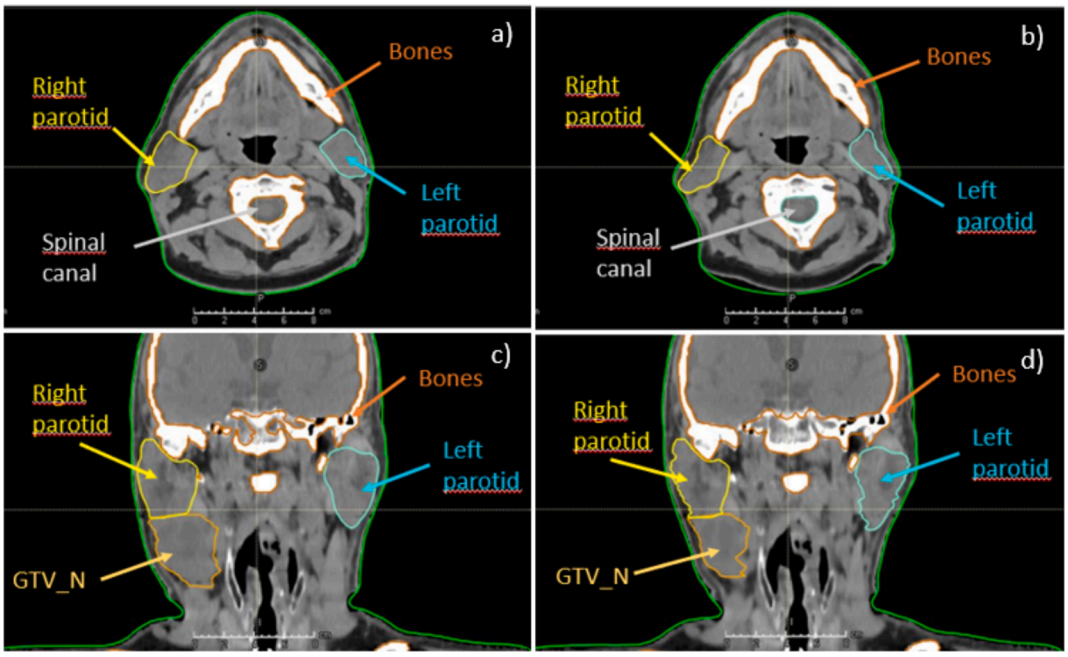

Fig. 1. Axial (a and b) and coronal (c and d) view of the hn_r phantom, with highlighted latero-cervical lymph node (GTV_N) and OARs. Pre (a and c) and post (b and d) deformation phantoms: the applied deformations led to clinical reliable volume reductions of 16%-17% for parotids and 26% for GTV_N.

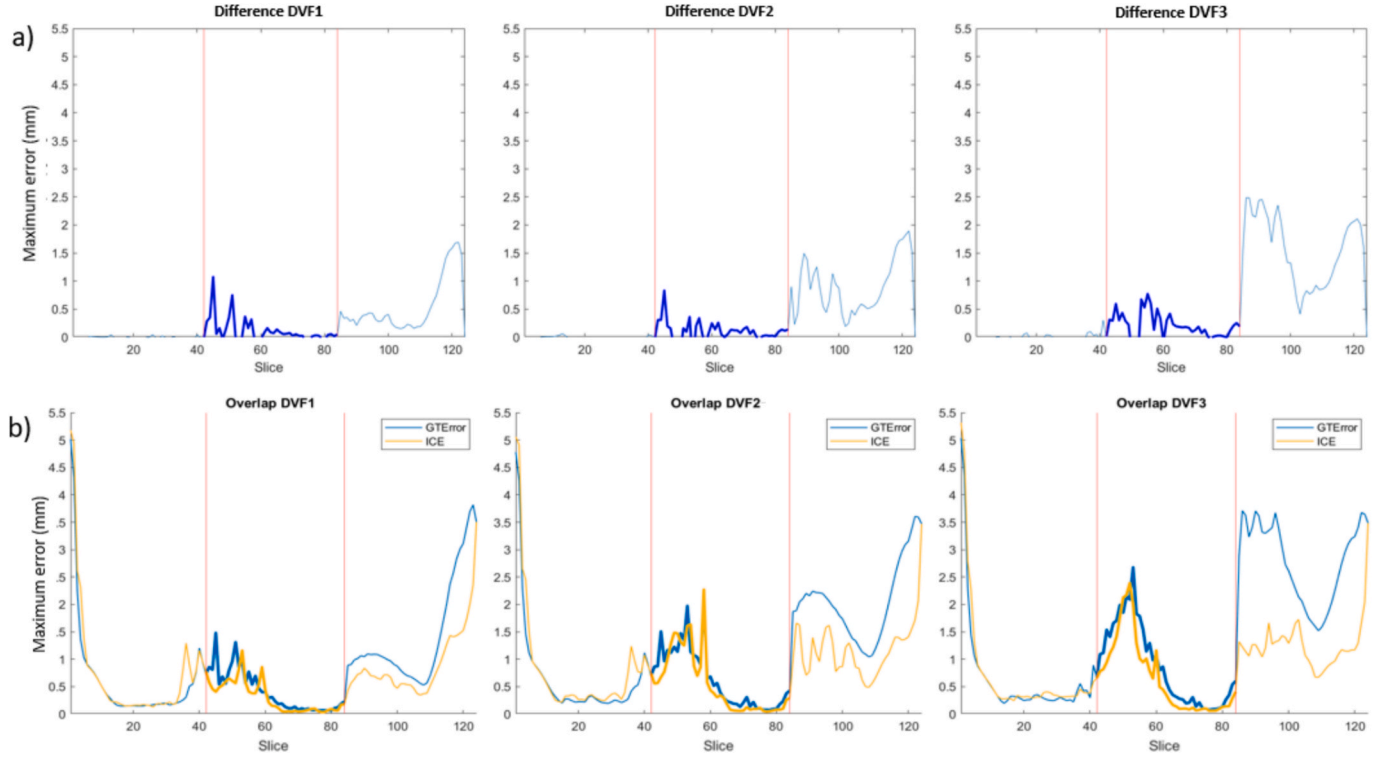

Fig. 2. Distributions of Maximum GTRE error and the corresponding Maximum ICE of the Raystation DIR algorithm calculated all over the slices of the HN_G phantom. The difference and the overlap between the two distributions are shown in panel a) and b) respectively. The vertical red bars highlight the region of clinical interest containing the contoured ROIs and the deformations applied by ImSimQA to generate the ground truth DVFs. DVF1-DVF3 denote the different magnitude of the applied deformations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

error scaling independently of anatomical complexity, providing a controlled baseline for magnitude relationships

4. Percentile-based correspondence (Fig. 5, HN_R). ICE and GTRE distributions were compared using matched percentiles to assess their relationship in the most anatomically realistic setting. The HN_R phantom was selected to evaluate ICE–GTRE correspondence under clinically plausible deformation patterns.

The potential influence of interpolation and resampling procedures on the resulting ICE maps was carefully considered. As discussed in Murr

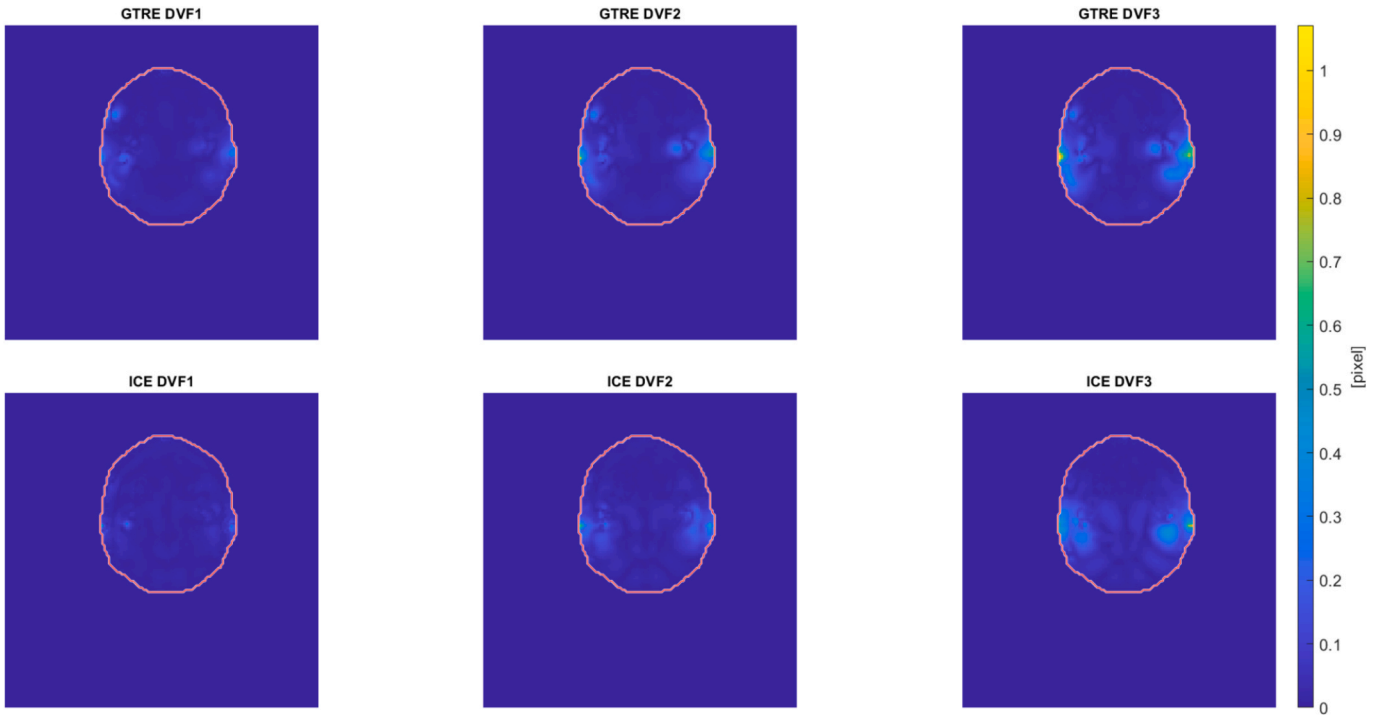

Fig. 3. GTRE (top) and ICE maps (bottom) of the Raystation DIR algorithm (ICERay) calculated on the HN_L phantom section corresponding to the maximum parotid shrinking for increasing intensities of applied deformations (units are expressed in pixels).

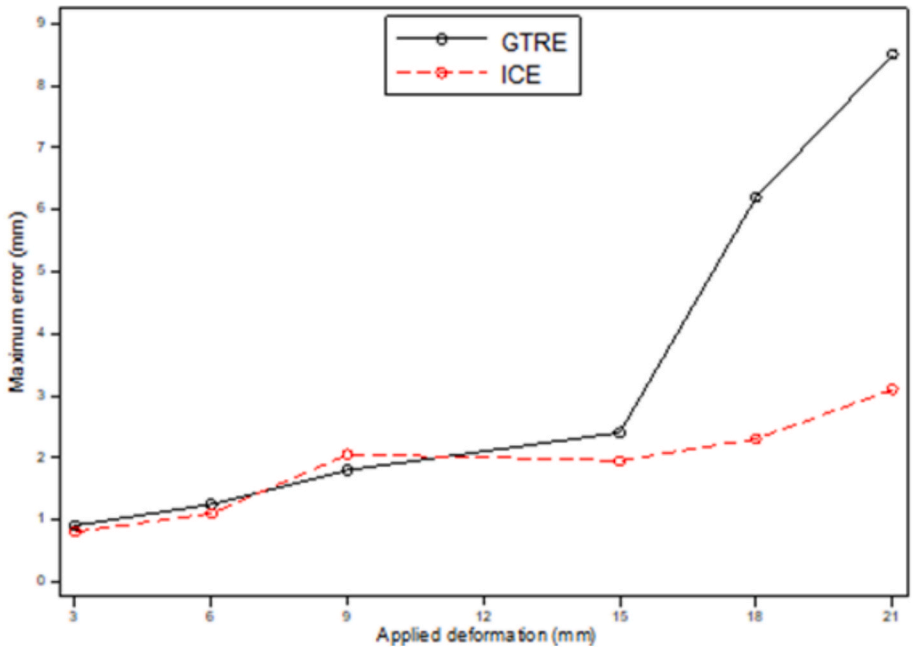

Fig. 4. Relationship between the GTRE and ICE of RayStation algorithm as a function of the applied deformation using geometric digital phantoms. The agreement between the two metrics is good up to to 15 mm deformation vector.

et al. (2023) [23], the choice of resampling strategy (direct dose mapping vs energy/mass transfer) and interpolation kernel can significantly affect voxel-wise results, introducing blurring or aliasing when multiple resampling steps are performed. To minimize these effects, a single "pull" interpolation step was applied to avoid multiple resampling operations.

2.4. Metrics and statistical analysis

Quantitative metrics were computed for both geometric and anatomical phantoms. ICE was calculated across the entire image volume, while classical reference metrics such as MDA and TRE were evaluated in the anatomical phantoms only. Target registration errors (TRE) were measured using 20 anatomical landmarks. Point-like and cuspid features such as vessel calcifications, neck muscles bifurcations and other suitable structures were identified within 2 cm from the PTV

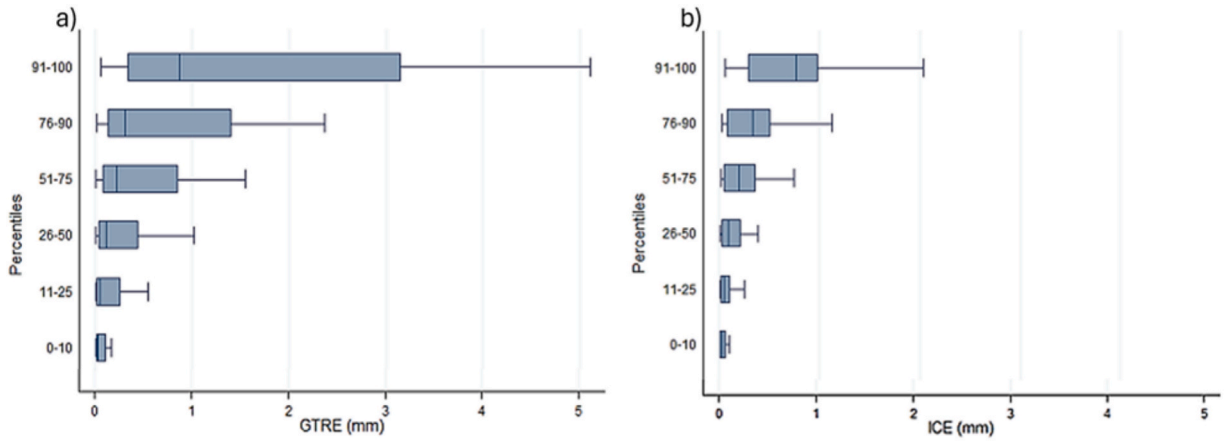

Fig. 5. Percentiles of the distribution of GTRE (panel a) and ICE_{Ray} (panel b) calculated all over 22 ROI volumes segmented on the HN phantoms.

borders by an expert radiation oncologist and crosschecked by a second operator. Localization uncertainties were estimated within 1.2 mm. Mean distance to agreement (MDA) was computed between reference and mapped ROIs for 22 structures (targets and OARs) in the HN phantoms.

Percentile analyses of ICE and GTRE were performed on the HN_R phantom, as it represents the most clinically realistic scenario with anatomically meaningful tumor and parotid volume changes. The relationship between ICE and both GTRE and TRE was evaluated using linear regression. Pearson correlation coefficients and p-values were reported. Finally the association between contour-based metrics (MDA) and voxel-wise errors (ICE/GTRE) was explored using the Kruskal-Wallis test.

Each metric was computed only in the scenarios where it is meaningful and technically applicable. Voxel-wise metrics (ICE and GTRE) were evaluated for all phantoms and deformation levels, whereas contour-based metrics (MDA) and landmark-based TRE were assessed only in the anatomical HN phantoms, where corresponding structures or anatomical landmarks exist. This avoids artificially extending metrics to cases where they cannot be defined or would not provide interpretable results.

ICE was expressed both in voxel units and milliliters, and its magnitude was interpreted relative to the voxel size of each dataset. The RayStation DIR algorithm was deemed locally inverse consistent when ICE was below the voxel dimension.

To evaluate ICE robustness, the extrema of GTRE and RayStation ICE values of the DIR mapped geometric phantoms were compared across all the aforementioned deformation levels (DEF1-DEF6).

All statistical analyses were conducted using Stata 14.2 (StataCorp, College Station, TX, USA). A p-value below 0.05 was deemed statistically significant.

3. Results

3.1. Validation of ground-truth DVFs

The ICE maps computed by ImSimQA generated DVFs showed negligible values throughout the image volume for both local and global transformations. In the latter case, boundary effects due to ThPs interpolant introduced minor discontinuities at the image edges (Supplementary Fig. E6). However, across the six increasing deformation levels (DEF1–DEF6, 3–21 mm), the maximum ICE in all datasets was consistently below 1 mm, which is well within the dataset voxel size, thereby validating the reliability of the ground-truth DVFs. The local transformations generated using CSRBF interpolation resulted in nearly zero ICE across the entire volume (Supplementary Materials, Figs. E8–E9), confirming the mathematical invertibility of CSRBF-based transformations on the whole image dataset.

Like geometric phantoms, ICE maps derived from anatomical HN phantoms exhibited near-zero values throughout the body contour, confirming the invertibility of ground truth DVFs. The maximum ICE values remained below the voxel size (Supplementary Material Fig E10 illustrates the slice-wise range of maximum ICE values for the HN_L phantom).

3.2. Relationship between the GTRE and ICE

The maximum_GTRE errors and ICE values obtained calculated throughout all over the slices of the HN_G digital phantoms are reported in Fig. 2.The red lines define the boundary of the region of clinical interest, including the ROIs segmented by the radiation oncologist and the domain of deformations applied by ImSimQA to generate the ground truth DVFs (DVF1, DVF2, DVF3). Within this region, the maximum GTRE was always below 3 mm, confirming the sub-voxel accuracy of the RayStation DIR algorithm. The ICE_{Ray} distribution closely mirrors the GTREs for the higher deformations (DVF2, DVF3, Fig. 2b): the positions of the maxima of the two error distributions are coincident showing a sub-millimetric agreement of the ICE metric with the local spatial registration error (Fig. 2a). For the lower deformation (DVF1) the trend is maintained, but there is a spatial mismatch between the two distributions; ICE metrics probably lack of sensitivity for small GTRE errors because of the invertibility and smoothing regularization terms contained in the algorithm implementation [22] as well as the partial volume effects associated to the deformation grid size (2.5 mm).

Fig. 3 illustrates the GTRE and ICE_{Ray} maps calculated for the HN_L phantom slice corresponding to the maximum shrinkage of the parotid glands. The errors shown at the top of the figure increase with the intensity of the applied deformations, primarily located around the parotids, reaching their highest values near the patient's external surface. It is worth noting that, even in the worst-case scenario (DVF 3), the maximum GTRE error was inferior to the pixel size, proving the robustness of the DIR RayStation algorithm to recover relevant local deformation with sub-voxel accuracy. The ICE_{Ray} maps at the bottom of the figure show spatial consistency with the GTRE distribution, indicating higher values in higher GTRE regions.

Regarding the robustness evaluation on the homogenous geometric phantom, ICE and GTRE agreed well up to 15 mm deformation (DEF4), where the DIR remained spatially accurate and ICE_{Ray} values stayed within voxel limits (Fig. 4). Beyond this point, ICE underestimated GTRE, reflecting the increasing dominance of regularization constraints in the DIR algorithm in the absence of anatomical features or ROI guidance.

The percentiles of the distribution of the GTREs and ICE_{Ray} calculated all over the ROI volumes of the HN_R phantoms are shown in

Fig. 5. The ICE_{Ray} appears systematically lower than GTRE with the median ± interquartile range (IQR) values of 0.8 ± 0.2 mm and 1.6 ± 0.4 mm respectively. It is worth noting that the DIR performed by Raystation demonstrates inverse consistency across the analyzed transformation domain (ICE < 2.5 mm). The upper decile of the GTRE distribution (Fig. 5a) shows outliers that exceed the sub-voxel (3 mm) tolerance requirement. These outliers were primarily found near the PTV ROIs in areas with challenging local deformations simulating non-mass preserving tumor response. This indicates that inverse consistency based on a predefined threshold, although necessary, is not sufficient to evaluate DIR accuracy.

Within the same clinically relevant error distribution interval, a significant linear relationship ($p < 0.001$) was observed between ICE and GTRE values (Fig. 6.a). The Pearson correlation coefficient for the linear regression was determined to be 0.852, indicating that the magnitude of ICE can be considered a reliable predictor of registration errors.

Switching the ground truth scenario to the set of point-like anatomical landmarks, a significant linear relationship ($p < 0.001$) was found between the ICE and TRE values (Fig. 6.b). The Pearson correlation coefficient was 0.683, suggesting a weaker correlation between the two variables compared to GTRE. However, the linear model predictions and the TREs showed strong agreement, with maximum differences smaller than the deformation grid size (2.5 mm), as illustrated by the plot of residuals from the linear fit in Fig. 6.c.

Regarding the relationship between GTRE, ICE and contour-based metrics, the MDA values were in a narrow range between 2.18 mm and 2.73 mm, an average of 2.47 ± 0.18 mm, thus below the voxel size assessing the good quality of the DIR ROI mapping. No significant associations ($p = 0.001$, Kruskal-Wallis test) were identified between MDA and the upper decile of both the GTRE and ICE distributions calculated within the corresponding volumes. This finding is anticipated and underscores that contour-based metrics are unable to predict the presence of critical ROI sub-volumes with errors greater than the image voxel size. This demonstrates the necessity for supplementary information, such as that provided by ICE maps, to evaluate the quality of DIR across the entire data set volume.

4. Discussion

This study evaluated inverse consistency error (ICE) as a voxel-wise metric for validating deformable image registration in radiotherapy. Using computational and anatomical phantoms, we found that ICE reflects spatial registration accuracy in feature-rich regions undergoing realistic anatomical changes, while its limitations become evident in homogeneous or large deformations.

Automated and quantitative methods for patient-specific DIR quality assurance are still lacking in adaptive workflows [22]. Classical contour-based metrics such as DSC and MDA [23,24] are widely used for evaluating DIR performance, but they cannot detect voxel-level inconsistencies and therefore provide limited insight into uncertainties relevant for dose mapping and accumulation [6–9,23–25]. This motivates the need for complementary voxel-wise metrics. Dose mapping is particularly sensitive to local DIR uncertainties [23,23,26–28], reinforcing the importance of voxel-wise metrics expressed in physical units. Existing DVF-based descriptors, such as Jacobian determinants, provide useful information but are not easily interpretable for routine QA.

ICE provides complementary information because it quantifies the mismatch between forward and backward DVFs [12,15,16] and can be computed directly from clinical registrations without manual input. The strong linear correlation observed between ICE and ground-truth registration error (GTRE) in the anatomical phantoms supports its potential as a voxel-level surrogate for local uncertainty in clinically relevant scenarios. The agreement with target registration error (TRE) derived from anatomical landmarks further confirms that ICE is sensitive to local discrepancies at spatial scales consistent with clinical localization uncertainties. As expected, no meaningful association was found between ICE/GTRE and contour-based metrics such as MDA, highlighting once more that ROI-level evaluations cannot detect sub-voxel local errors. [29–32].

These findings were consistent across both ROI-based and landmark-based evaluations, confirming previously reported associations between inverse consistency and spatial registration error.

The robustness of ICE as a surrogate for spatial DIR uncertainty was also explored using homogeneous geometric phantoms across multiple deformation levels. This analysis revealed that there was a threshold beyond which ICE was no longer a reliable predictor of registration errors. Specifically, for local deformations exceeding 15 mm, the RayStation algorithm remained inverse-consistent (ICE < 2.5 mm), but diverged significantly from GTRE. Under these conditions the DVF internal mathematical consistency is prioritized on the voxel intensity metric and the regularization ROI constraints leading to the well-known observation that inverse consistency is necessary but not sufficient for spatial accuracy assessment [14]. This is the most significant drawback of the metrics and highlight that ICE sensitivity to registration errors can depend from the DIR algorithm implementation and image quality. In such clinical situations, yet to be explored, a multilayered strategy including a crosscheck with differential based metrics (Jacobian) can be used as complement to enforce the detection of not physically plausible DVFs.

Our study has several strengths, including the use of controlled phantoms with known ground truth, the evaluation of multiple deformation strategies, and the explicit comparison of ICE with GTRE, TRE, and MDA. The workflow used to compute ICE and GTRE reflects realistic

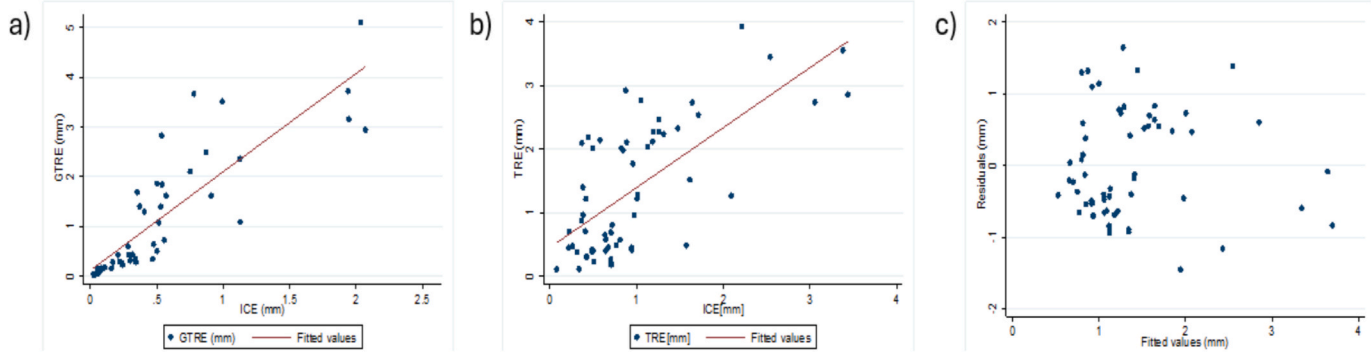

Fig. 6. Graphs illustrating the results of a linear regression between ICE_{Ray}, GTRE and TRE:a) A strong correlation between ICE_{Ray} and GTREs ($R = 0.852$) was found. The analysis was performed on the upper deciles of error distributions considered clinically relevant. b) A significant correlation is here reported between ICE_{Ray} and TRE ($p < 0.001$, $R = 0.683$); c) The plot of residuals versus the fitted values indicated that the difference between the linear fit prediction and the TREs was less than 2 mm.

clinical handling of DVFs, including interpolation, resampling, and resolution harmonization, which are known to influence voxel-wise accuracy [23].

Some limitations must also be acknowledged. First, the analysis was conducted on a single commercial DIR algorithm, and the sensitivity of ICE may vary across platforms with different regularization schemes and optimization strategies. Second, the phantom-based evaluation focused on head and neck anatomy; scenarios involving sliding organs, high-motion regions, or complex internal deformations were not included. Finally, while ICE provides valuable voxel-wise information, it should not be interpreted as a stand-alone guarantee of anatomical correctness in regions with tissue regression or emergence, where the assumption of one-to-one correspondence is violated.

Overall, ICE offers a practical, automated, and interpretable method for assessing DIR uncertainty at the voxel level. When applied within its validity domain, ICE can complement existing QA methods, support more reliable dose deformation workflows, and strengthen decision-making in adaptive and re-irradiation radiotherapy. ICE maps overlaid on patient anatomy, along with ICE Volume Histograms—calculated like dose-volume histograms by binning ICE values within the ROI—may offer a practical approach for enhancing patient-specific QA in clinical workflows (Fig. E14).

CRediT authorship contribution statement

Gianfranco Loi: Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Marco Fusella:** Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Stefania Zara:** Methodology, Writing – review & editing. **Marica Vagni:** Methodology, Writing – review & editing, Writing – original draft. **Nicola Michielli:** Methodology, Writing – review & editing. **Orlando Zaccaria:** Methodology, Writing – review & editing. **Lorenzo Placidi:** Methodology, Writing – review & editing. **Pierfrancesco Franco:** Methodology, Writing – review & editing. **Filippo Molinari:** Methodology, Writing – review & editing. **Christian Fiandra:** Conceptualization, Methodology, Formal analysis, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: SZ, OZ are employers of Tecnologie Avanzate Srl, Italian distributor of TPS RaySearch RayStation. All other authors don't have known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper..

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2026.100916.

References

[1] Castillo R, Castillo E, Guerra R, et al. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys Med Biol 2009;54:1849–70. https://doi.org/10.1088/0031-9155/54/7/001.

[2] Loi G, Fusella M, Roggio A, et al. Performance of commercially available deformable image registration platforms for contour propagation using patient-based computational phantoms: a multi-institutional study. Med Phys 2018;45:748–57. https://doi.org/10.1002/mp.12737.

[3] Kim H, Ko S, Kazanzides P, Fichtinger G, Liao R. Quantitative analysis tools and digital phantoms for deformable image registration quality assurance. Technol Cancer Res Treat 2015;14:428–39. https://doi.org/10.1177/1533034614553891.

[4] Castillo R, Pham N, Ansari S, et al. A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. Phys Med Biol 2013;58:2861–78. https://doi.org/10.1088/0031-9155/58/9/2861.

[5] Brock KK. Results of a multi-institution deformable registration accuracy study (MIDRAS). Int J Radiat Oncol Biol Phys 2010;76:583–96. https://doi.org/10.1016/j.ijrobp.2009.06.031.

[6] Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: report of AAPM TG-132. Med Phys 2017;44:e43–76. https://doi.org/10.1002/mp.12256.

[7] Schultheiss TE, Tomé WA, Ayan AS. It is not appropriate to deform dose along with deformable image registration in adaptive radiotherapy. Med Phys 2012;39:6531–3. https://doi.org/10.1118/1.4722968.

[8] Juang T, Das S, Adamovics J, et al. On the need for comprehensive validation of deformable image registration, investigated with a novel 3-dimensional deformable dosimeter. Int J Radiat Oncol Biol Phys 2013;87:414–21. https://doi.org/10.1016/j.ijrobp.2013.05.045.

[9] Velec M, Moseley J, Eccles C, et al. Utility and validation of biomechanical deformable image registration in low-contrast images. Pract Radiat Oncol 2015;5:e401–8. https://doi.org/10.1016/j.prro.2015.01.011.

[10] Yanovsky I, et al. Unbiased volumetric registration via nonlinear elastic regularization. Available at: HAL Research Archive 2008. https://hal.inria.fr/inria-0062976.

[11] Jurkovic IA, et al. Objective assessment of the quality and accuracy of deformable image registration. J Med Phys 2020;45:156–67. https://doi.org/10.4103/jmp.JMP_47_19.

[12] Christensen GE, Johnson HJ. Consistent image registration. IEEE Trans Med Imaging 2001;20:568–82. https://doi.org/10.1109/42.932742.

[13] Yang D, Li H, Low DA, Deasy JO, Palta JR. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. Phys Med Biol 2008;53:6143–61. https://doi.org/10.1088/0031-9155/53/21/005.

[14] Loi G, et al. CT-to-CBCT deformable image registration for contour propagation using head and neck computational phantoms: a multicenter study. Pract Radiat Oncol 2020;10:125–32. https://doi.org/10.1016/j.prro.2019.11.011.

[15] Bender ET, Tomé WA. The utilization of consistency metrics for error analysis in deformable image registration. Phys Med Biol 2009;54:5561–77. https://doi.org/10.1088/0031-9155/54/18/014.

[16] García-Mollá R, et al. Validation of a deformable image registration produced by a commercial TPS in head and neck. Phys Med 2015;31:219–23. https://doi.org/10.1016/j.ejmp.2015.01.007.

[17] Wang H, Dong L, O'Daniel J, et al. Implementation and validation of a 3D deformable registration algorithm for targeted prostate cancer radiotherapy. Int J Radiat Oncol Biol Phys 2005;61:725–35. https://doi.org/10.1016/j.ijrobp.2004.07.677.

[18] Bookstein FL. Principal warps: thin-plate splines and the decomposition of deformations. IEEE Trans Pattern Anal Mach Intell 1989;11:567–85.

[19] Wendland H. Piecewise polynomial, positive definite and compactly supported radial functions for soft tissue deformation. In: Proc IEEE Int Symp Biomed Imaging; 2004. p. xxx–xxx.

[20] Johnson HJ, Christensen GE. Consistent landmark and intensity-based image registration. IEEE Trans Med Imaging 2002;21:450–61. https://doi.org/10.1109/TMI.2002.1009381.

[21] Wachowiak MP, et al. Compact support radial basis functions for soft tissue deformation. In: Proc IEEE Int Symp Biomed Imaging; 2004. p. xxx–xxx.

[22] Weistrand O, Svensson S. The ANACONDA algorithm for deformable image registration in radiotherapy. Med Phys 2015;42:40–53. https://doi.org/10.1118/1.4894702.

[23] Murr M, Brock KK, Fusella M, et al. Applicability and usage of dose mapping/accumulation in radiotherapy. Radiother Oncol 2023;182:109527. https://doi.org/10.1016/j.radonc.2023.109527.

[24] Kubota Y, et al. Evaluation of intensity- and contour-based DIR accuracy in pancreatic cancer patients. Cancers 2019;11:1447. https://doi.org/10.3390/cancers11101447.

[25] Thiong'o CM, Davey A, Appelt A, et al. Dose mapping using image registration for reirradiation: a systematic review. Int J Radiat Oncol Biol Phys 2025;in press. https://doi.org/10.1016/j.ijrobp.2025.09.053.

[26] Salguero FJ, et al. Estimation of 3D intrinsic dosimetric uncertainties resulting from using DIR for dose mapping. Med Phys 2011;38:343–53. https://doi.org/10.1118/1.3528201.

[27] Murphy MJ, et al. A method to estimate the effect of DIR uncertainties on daily dose mapping. Med Phys 2012;39:573–80. https://doi.org/10.1118/1.3673772.

[28] Qin A, et al. Technical note: the impact of deformable image registration methods on dose warping. Med Phys 2018;45:1287–94. https://doi.org/10.1002/mp.12741.

[29] Nenoff L, Amstutz F, Murr M, et al. Review and recommendations on deformable image registration uncertainties for radiotherapy applications. Phys Med Biol 2023;68:24TR01. https://doi.org/10.1088/1361-6560/ad0d8a.

[30] Bosma LS, Hussein M, Jameson MG, et al. Tools and recommendations for commissioning and QA of deformable image registration in radiotherapy. Phys Imaging Radiat Oncol 2024;32:100647. https://doi.org/10.1016/j.phro.2024.100647.

[31] Saleh Z, Apte A, Sharp G, et al. The distance discordance metric: a novel approach to quantifying spatial uncertainties in deformable image registration. Phys Med Biol 2014;59:733–46. https://doi.org/10.1088/0031-9155/59/3/733.

[32] Scaggion A, Fiandra C, Loi G, Vecchi C, Fusella M. Free-to-use DIR solutions in radiotherapy: benchmark against commercial platforms through a contour-propagation study. Phys Med 2020;74:110–7. https://doi.org/10.1016/j.ejmp.2020.05.011.