Penalized least-square CT reconstruction without and with statistical weights: effect on lesion detection performance with human observers

V. Haase, A. Griffith, Z. Guo, M. B. Oktay, K. Hahn, H. Schoendube, K. Stierstorfer, F. Noo

***Abstract*—Iterative CT reconstruction with the penalized least-square model may offer significant gains in terms of image quality at equal dose, and may thereby allow either dose reduction or improved diagnostic. In this work, we are interested in evaluating image quality improvements that result from using statistical weights in this model. Image quality is assessed in terms of lesion detection with unknown location, using the principles of LROC analysis with human observers. Reconstruction without and with statistical weights are compared for two penalties: a quadratic penalty, and an edge-preserving penalty. Interestingly, our study failed showing any major improvements due to the use of weights. Furthermore, it was even observed that performance with weights could be even worse, possibly due to the utilization of weights leading to disturbing discretization errors. Because there are a lot of degrees of freedom in our experimental set-up, it should not be concluded that statistical weights are not useful. However, we can state that improvements are not straightforward and may depend on many aspects including the task and anatomical location. This observation is valuable from a computational viewpoint since using statistical weights generally leads to long reconstruction times; if weights can be ignored in some settings, reconstruction times can be largely improved for these settings.**

I. Introduction

Iterative CT reconstruction using advanced statistical models is currently a hot topic of research. For diagnostic CT imaging, the main advantage that is sought from such reconstruction is improved image quality, which may either be used to augment the role of CT in medicine and clinical research or to reduce dose to the patient for conventional CT scans. Encouraging clinical results have been published [1]–[5].

Among a number of options, the penalized least-square model with statistical weights [6], [7] is highly popular. In this model, the statistical weights represent the variance of the measurements. These weights are applied in the data fidelity term to enable accounting for different noise levels across measurements. To regularize the reconstruction, the data fidelity term is further balanced with a penalty term that typically constrains differences between neighboring voxel values.

Designing an efficient iterative algorithm to solve the penalized weighted least-square reconstruction problem turns out to be highly difficult, particularly for convergence within 1 HU from the desired solution. The wide dynamic range of statistical weights is largely responsible for this situation. In this work, we are interested in evaluating the image quality improvement brought by the statistical weights, in comparison with using the same model but without weights. The evaluation is carried out in fanbeam geometry, and task-based assessment of image quality is employed with human observers.

The paper is organized as follows. First, we present our evaluation methodology. Next, we show our main results. Afterwards, we offer a brief discussion and conclusions.

V. Haase, K. Hahn, H. Schoendube and K. Stierstorfer are with Siemens Healthcare GmbH, Forchheim, Germany. Z. Guo, M. B. Oktay and F. Noo are with the Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, Utah, USA. The concepts presented in this paper are based on research and are not commercially available. This work was partially supported by Siemens Medical Solutions, USA.

II. Evaluation methodology

A. Problem formulation

Let $\underline{x}$ be the vector of image pixel values to be reconstructed and let $\underline{b}$ be the vector of CT measurements. The desired reconstruction is defined as the minimizer of the following objective function with positivity constraint on the entries of $\underline{x}$:

$$\Phi(\underline{x},\underline{b}) = \Phi_1(\underline{x},\underline{b}) + \beta\,\Phi_2(\underline{x}) \tag{1}$$

where $\Phi_1(\underline{x},\underline{b})$ is the data fidelity term, and $\Phi_2(\underline{x})$ is the penalty (regularization) term. The balance between these two terms is controlled by parameter $\beta > 0$. The expression for the data fidelity term is

$$\Phi_1(\underline{x},\underline{b}) = \left\|\mathbf{W}^{-1/2}(\mathbf{A}\underline{x}-\underline{b})\right\|^2 \tag{2}$$

where $\mathbf{A}$ is a forward projection matrix, and $\mathbf{W}$ is a diagonal matrix with entries equal to the exact variance of the measurements. When used, the statistical weights were normalized so that the mean value for the central ray is equal to one, when taking the mean value over the views. The expression for the regularization term is

$$\Phi_2(\underline{x}) = \frac{1}{2}\sum_i\sum_j \omega_{ij}\psi(x_i - x_j) \tag{3}$$

with $\omega_{ij} = 1$ for horizontal and vertical neighbor pixels, $\omega_{ij} = 1/\sqrt{2}$ for diagonal neighbor pixels, and $\omega_{ij} = 0$ otherwise. Two choices are considered for $\psi(t)$: (i) $\psi(t) = t^2$, which amounts to performing reconstruction with FBP-like (quadratic) regularization, and (ii)

$$\psi(t) = \delta\cdot[|t/\delta| - \log(1+|t/\delta|)] \tag{4}$$

which preserves edges through parameter δ.

B. Iterative reconstruction technique

Given the chosen expressions for $\Phi_1(\underline{x},\underline{b})$ and $\Phi_2(\underline{x})$, the desired reconstruction can be robustly computed using the iterative coordinate descent (ICD) method [8]–[11]. We have employed this method. More precisely, the objective function was sequentially minimized by iterating over each pixel value, one at a time, in a random order. The result obtained after iterating over all pixels values is called one iteration. The last

focal spot size	0.12 cm × 0.09 cm
anode angle	7°
source trajectory radius	57 cm
source to detector distance	104 cm
number of detector elements	672
angular detector width	0.001231 radians
detector row height	0.128 cm
number of projections per turn	1160
incoming number of photons per ray:	
for quadratic penalty	180,000
for edge-preserving penalty	90,000

TABLE I
DATA SIMULATION PARAMETERS.

iteration was chosen as that for which all pixel values changed by less than $0.0001/\text{cm}$. All iterations were initiated with a zero image to prevent any bias from an initial reconstruction.

C. Data simulation

Reconstructions were performed from computer-simulated fanbeam data of the FORBILD head phantom. The geometry of 3rd generation CT was used, with parameters given in Table 1. In order to model shift-variant resolution effects, the X-ray data simulation included a sub-sampling of the X-ray tube focal spot, of each detector, and of each source position. Each CT measurement was thus formed as an average of 48 line integrals, with the average taken before applying the logarithm, and with each line integral computed from analytical expressions.

Several noisy data sets were created to simulate repeated scans. The noise was based on Poisson statistic using a realistic body-size Gaussian-shaped bowtie filter. The beam was assumed to be monochromatic. A low energy of 40kV was used to accentue variations in noise level across the CT measurements, and thus the effect of statistical weights.

D. Image reconstruction

The image was viewed as a set of samples on a Cartesian grid of 551×701 pixels with a uniform sampling distance of 0.0375 cm in both directions. Matrix A was formed using the distance-driven method [12], which accounts for the finite width of the detector pixels.

Each reconstruction was split into two sub-images, one corresponding to the upper half of the phantom, and the other one to the lower half. The two sub-images are seen as independent cases for the human observer study. Stastistical independence was ensured by discarding a central strip of 15 pixels. Each sub-image either included no lesion or exactly one lesion, which was included during data simulation. When present, the lesion was always within the low-contrast brain-tissue region of the phantom, with no overlap with the bones. The lesion was a 7 mm diameter disk with random contrast varying between 20 and 30 HU.

E. Task-based assessment of image quality

Image quality was assessed in terms of lesion detectability with unknown location, using the principles of LROC analysis. The area under the LROC curve, called AUC, was employed as figure-of-merit. This AUC value was directly evaluated using an alternative forced choice experiment that involved two human observers (readers) for reconstructions with quadratic penalty, and three human observers for reconstructions with edge-preserving penalty.

Each observer participated in two sessions, within each of which the observer evaluated images reconstructed without and with statistical weights. Each session amounted to first reading training and then testing cases for one reconstruction method, and next, reading training and testing cases for the second reconstruction method. The number of training cases was always 40, and the number of testing cases was always 160, so that each session included 400 cases. Random numbers were used to decide which reconstruction method each session would start with. These numbers changed from observer to observer, as well as from session to session. All observers read the cases in the same dimmed room (10 lux), on a medical-grate monitor that was calibrated according to the ACR Technical Standard for Electronic Practice. To maximize statistical power, the cases were statistically independent of each other from one session to another, as well as from one observer to another, and the same repeated scans were used for both reconstruction methods. In other words, the cases were paired across methods, and unpaired across sessions and readers. Note also that, in our statistical analysis, we view the readers as fixed effects; this setting is deemed satisfactory given the high level of control associated with our studies and is indirectly supported by the observation that all readers perform fairly similarly.

III. RESULTS

A. Visual impression

Figures 1 and 2 show negative and positive cases for reconstruction with quadratic penalty. The cases in figure 1 are obtained without statistical weights, whereas the cases in figure 2 are obtained with statistical weights. The data sets are the same for both figures. Some major differences can be observed in terms of noise correlations. The reconstruction without weight displays more noise streaks. However, the reconstruction with weights exhibit more discretization errors (around the right ear as well as around the air cavity).

Figures 3 and 4 show negative and positive cases for reconstruction with edge-preserving penalty. The cases in figure 3 are obtained without statistical weights, whereas the cases in figure 4 are obtained with statistical weights. As before, the data sets are the same for both figures. Compared to reconstruction with quadratic penalty, we note that reconstruction without weights still displays more noise streaks, but reconstruction with weights does not exhibit the disturbing discretization errors that were previously observed.

B. Detectability performance

Figure 5 displays the main results obtained with the LROC studies, for both reconstruction with quadratic penalty (left column) and reconstruction with edge-preserving penalty (right column). In each column, the error bars for the top and middle plots correspond to individual 95% confidence intervals, whereas the error bars for the bottom plot correspond to

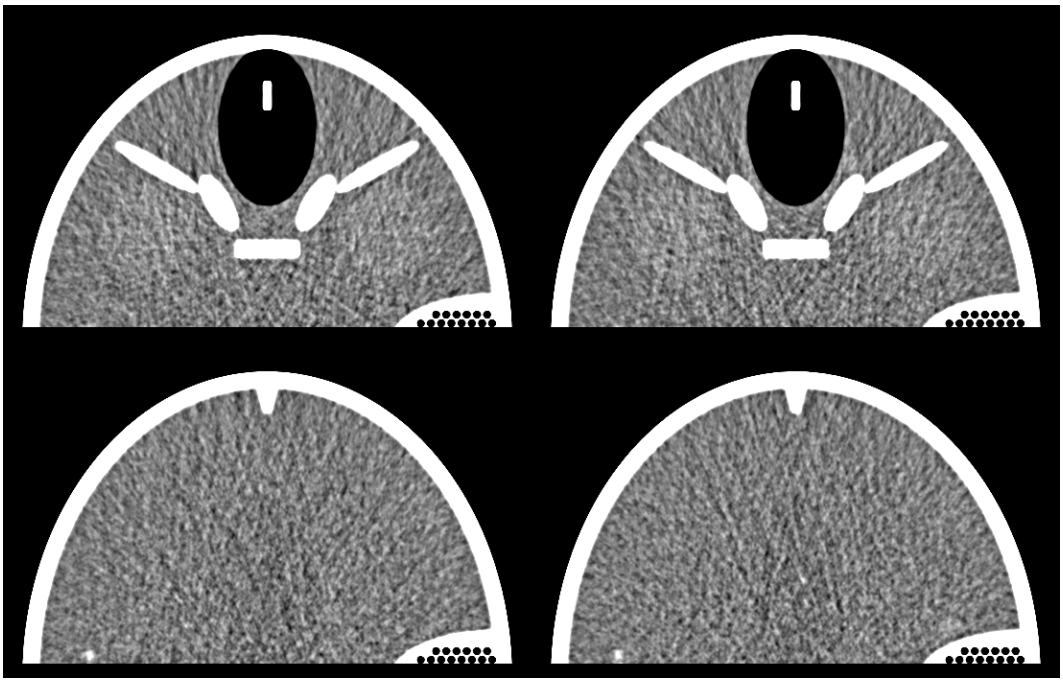

Fig. 1. Examples of reconstructions with quadratic penalty without statistical weigths. (left column) Lesion-free cases. (right column) Cases with exactly one lesion present. Grayscale: $[-30, 130]$ HU.

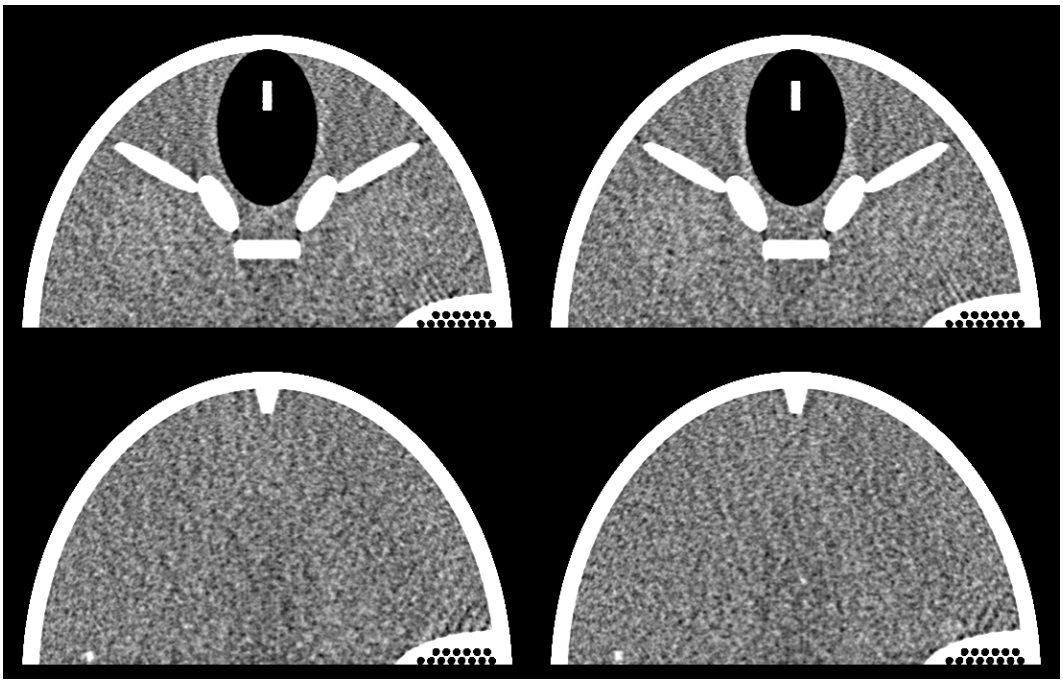

Fig. 2. Examples of reconstructions with quadratic penalty and with statistical weigths. (left column) Lesion-free cases. (right column) Cases with exactly one lesion present. Grayscale: $[-30, 130]$ HU.

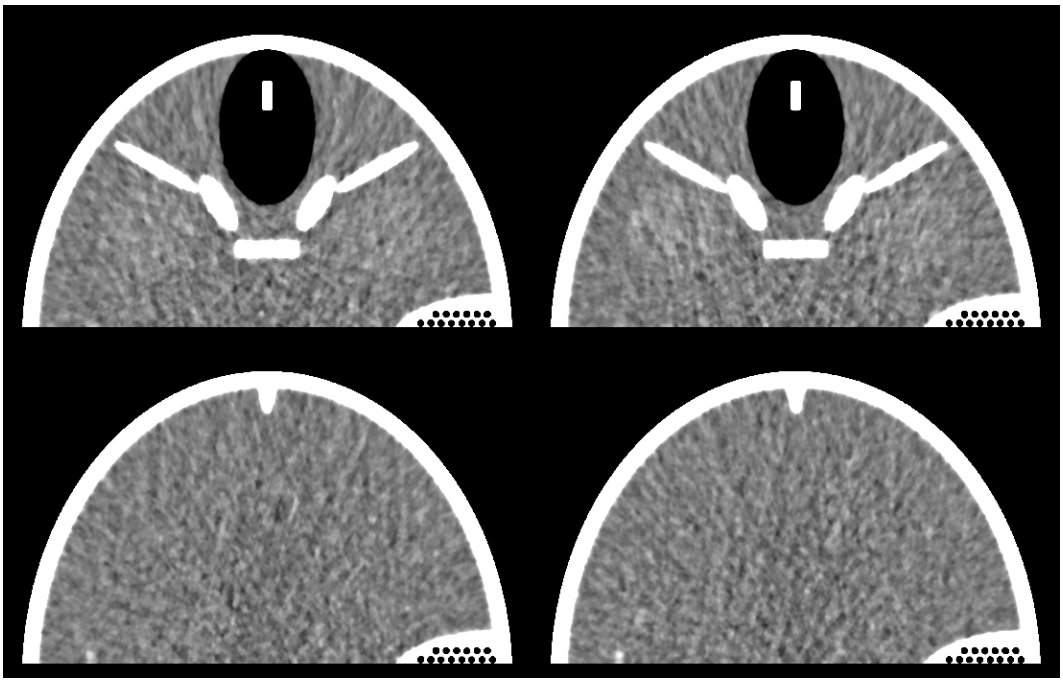

Fig. 3. Examples of reconstructions with edge-preserving penalty without statistical weigths. (left column) Lesion-free cases. (right column) Cases with exactly one lesion present. Grayscale: $[-30, 130]$ HU.

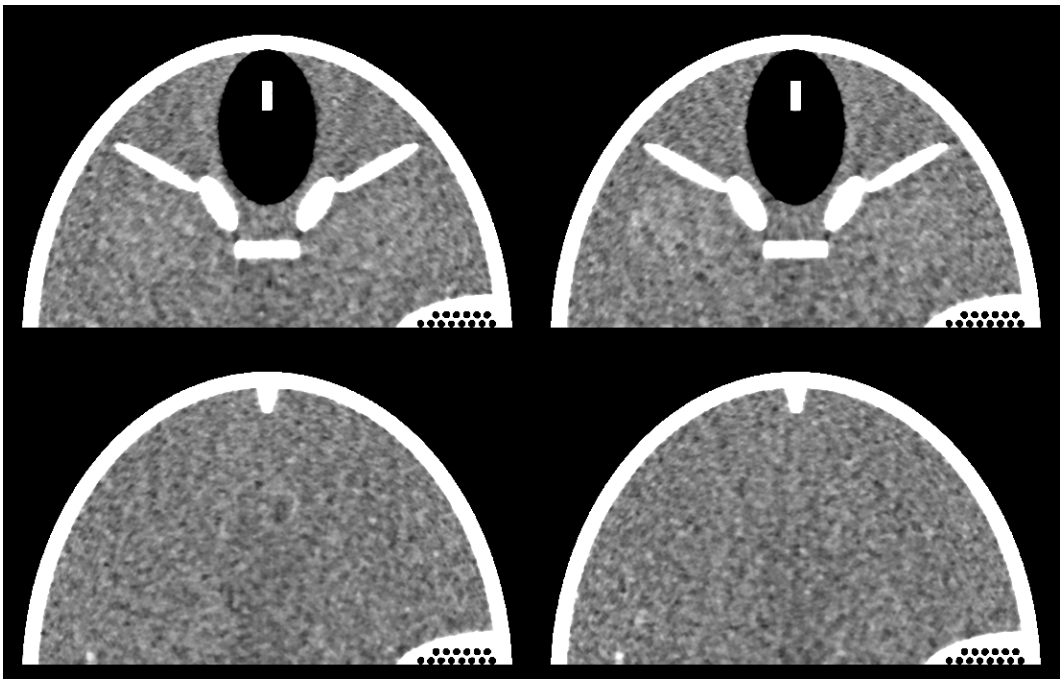

Fig. 4. Examples of reconstructions with edge-preserving penalty and with statistical weigths. (left column) Lesion-free cases. (right column) Cases with exactly one lesion present. Grayscale: $[-30, 130]$ HU.

joint 95% confidence intervals (based on Bonferoni inequality). For reconstruction with quadratic penalty, we observe that the reader-and-session-averaged performance is slightly worse for reconstruction with statistical weights than for reconstruction without weights. For reconstruction with edge-preserving penalty, we observe that the reader-and-session-averaged performance is, with 95% confidence, essentially the same for both reconstruction without and with statistical weights. Note that the size of the error bars in the bottom plots indicate that the statistical accuracy for these observations is fairly strong. Note also that the observer performance variation from one session to another is consistent with the size of the error bars; this variation is associated with the effect that cases in one session can easily be globally more challenging than cases in another session.

IV. Discussion and Conclusion

In this work, we reported results of two LROC studies with human observers. These studies aimed at evaluating improvements resulting from the use of statistical weigths in penalized least-square CT reconstruction. Interestingly, the study failed showing any major improvements due to the use of weights. Furthermore, it was even observed that performance with weights could be even worse, possibly due to the utilization of weights leading to disturbing discretization errors. Because there are a lot of degrees of freedom in our experimental set-up, it should not be concluded that statistical weights are not useful. However, we can state that improvements are not straightforward and may depend on many aspects including the task and anatomical location. This observation is valuable from a computational viewpoint since using statistical weights generally leads to long reconstruction times; if weights can be ignored in some settings, reconstruction times can be largely improved for these settings.

Many important questions remain open, two of which are as follows. First, it may be that statistical weights play a larger role when lesion detectability has to be performed with human anatomical variations. Second, it may be that statistical weights play a larger role when parameter β in the objective function is replaced by a voxel-dependent weight defined from the statistical weigths, as suggested in [13]. We are interested in investigating these two questions in the future.

References

[1] Pickhardt, P. J., Lubner, M. G., Kim, D. H., Tang, J., Ruma, J. A., del Rio, A. M., and Chen, G.-H., "Abdominal CT with model-based iterative reconstruction (mbir): Initial results of aprospective trial comparing ultralow-dose with standard-dose imaging," *AJR: American Journal of Roentgenology* **199**(6), 1266–74 (2012).

[2] Desai, G. S., Uppot, R. N., Yu, E. W., Kambadakone, A. R., and Sahani, D. V., "Impact of iterative reconstruction on image quality and radiation dose in multidetector CT of large body size adults," *European Radiology* **22**(8), 1631–1640 (2012).

[3] Neroladaki, A., Botsikas, D., Boudabbous, S., Becker, C. D., and Montet, X., "Computed tomography of the chest with model-based iterative reconstruction using a radiation exposure similar to chest X-ray examination: preliminary observations," *European Radiology* **23**(2), 360–366 (2013).

[4] Vardhanabhuti, V., Olubaniyi, B., Loader, R., Riordan, R. D., Williams, M. P., and Roobottom, C. A., "Image Quality Assessment in Torso Phantom Comparing Effects of Varying Automatic Current Modulation with Filtered Back Projection, Adaptive Statistical, and Model-Based Iterative Reconstruction Techniques in CT," *Journal of Medical Imaging and Radiation Sciences* **43**(4), 228–238 (2012).

[5] Hara, A. K., Paden, R. G., Silva, A. C., Kujak, J. L., Lawder, H. J., and Pavlicek, W., "Iterative Reconstruction Technique for Reducing Body Radiation Dose at CT: Feasibility Study," *AJR* **193**, 764–771 (2009).

[6] Fessler, J. A., "Penalized Weighted Least-Squares Image Reconstruction for Positron Emission Tomography," *IEEE Transaction on Medical Imaging* **13**(2), 290–300 (1994).

[7] Wang, J., Li, T., Lu, H., and Liang, Z., "Penalized Weighted Least-Squares Approach to Sinogram Noise Reduction and Image Reconstruction for Low-Dose X-Ray Computed Tomography," *IEEE Transactions on Medical Imaging* **25**(10), 1272–1283 (2006).

[8] Abatzoglou, T. and O'Donnell, B., "Minimization by coordinate descent," *J. of Opt. Theory Appl.* **36**(2), 163–174 (1982).

[9] Luo, Z. Q. and Tseng, P., "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Opt. Theory Appl.* **72**(1), 7–35 (1992).

[10] Thibault, J.-B., Sauer, K., Bouman, C., and Hsieh, J., "A three-dimensional statistical approach to improve image quality for multi-slice helical CT," *Med. Phys.* **34**(11), 4526–44 (2007).

[11] Yu, Z., Thibault, J.-B., Bouman, C. A., Sauer, K. D., and Hsieh, J., "Fast Model-Based X-Ray CT Reconstruction Using Spatially Nonhomogeneous ICD Optimization," *IEEE Transaction on Image Processing* **20**(1), 161–175 (2011).

[12] De Man, B. and Basu, S., "Distance-driven projection and backprojection in three dimensions," *Physics in Medicine and Biology* **49**, 2463 – 2474 (2004).

[13] Cho, J. H. and Fessler, J. A., "Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3-d x-ray CT," *IEEE transactions on medical imaging* **34**(2), 678–689 (2015).

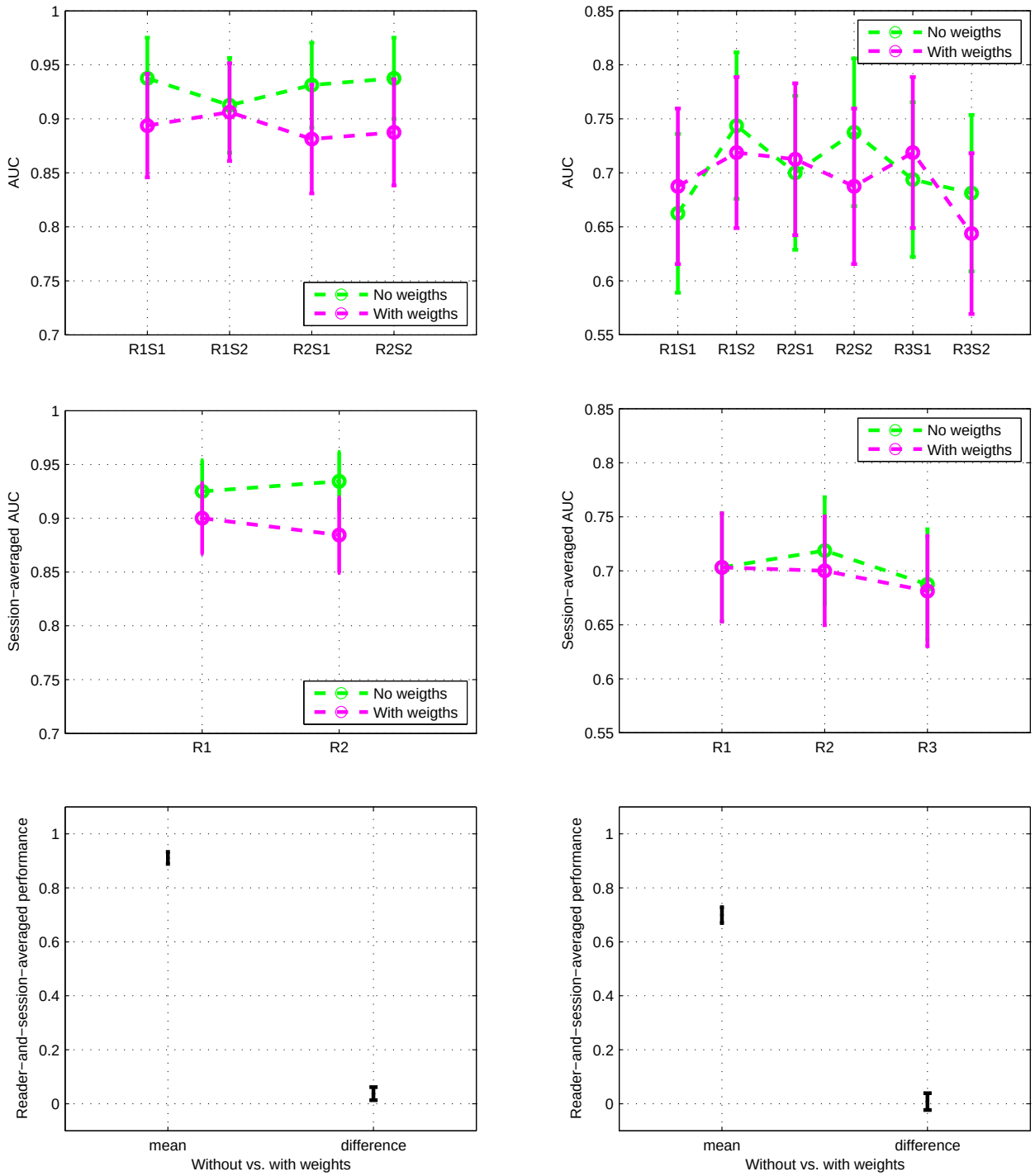

Fig. 5. Results from the LROC studies for reconstruction with quadratic penalty (left column) and with edge-preserving penalty (right column). In each column: (top) AUC value obtained for each observer and each session, (middle) AUC value obtained for each observer after average over sessions, (bottom) difference (mean, resp.) in reader-and-session-averaged AUC between (over, resp.) reconstruction without weights and reconstruction with weights.