Incorporating prior information in a no-gold-standard technique to assess quantitative SPECT reconstruction methods

Abhinav K. Jha and Eric C. Frey
Division of Medical Imaging Physics, Department of Radiology and Radiological Sciences,
Johns Hopkins University, Baltimore, MD, USA

***Abstract*—There is a need for objective evaluation methods to assess systems and methods being developed for quantitative single-photon emission computed tomography (SPECT) imaging when the true value of the quantitative parameter is not known, as is the case in patient studies. We have developed a no-gold-standard technique for assessing reconstruction methods for SPECT when the end task is estimating the activity concentration within a known volume of interest (VOI). The technique posits a linear relationship between the true and estimated mean activity concentrations and estimates the parameters of this linear relationship from the measured mean activity concentrations using a maximum-likelihood (ML) method. We have demonstrated the performance of this technique in ranking reconstruction methods for quantitative SPECT using realistic simulation studies. However, to rank the methods accurately, the technique requires a large number of patient studies, which are often not available. In this regard, prior information about a subset of the parameters of the linear relationship can be obtained from experimental phantom studies. To incorporate this information, we pose the no-gold-standard method as a maximum-a-posteriori (MAP) estimation problem. We devise statistical models for the prior information about the parameters and include these models in the MAP-estimation framework. A study of the performance of the proposed MAP-based no-gold-standard method with simulation studies demonstrated its ability to yield accurate rankings, as quantified using the Spearman's rank correlation coefficent, with even a small number of patient studies. The method also consistently outperformed the ML-based no-gold-standard technique.**

Keywords: No-gold-standard methods, quantitative SPECT, Objective evaluation of reconstruction methods.

I. Introduction

Quantitative single-photon emission computed tomography (SPECT) is emerging as an important tool for clinical studies and biomedical research [1], [2]. In quantitative SPECT, the end task is estimating a certain functional or anatomical parameter about the object of interest from a SPECT image, such as quantifying the mean activity concentration within a certain volume of interest (VOI) from a patient image. Many imaging systems and methods are being designed to provide accurate quantitative SPECT imaging. There is a need for objective methods to optimize and evaluate these systems and algorithms on the basis of their performance in the task of estimating the parameter of interest accurately and precisely. Such objective evaluation is greatly simplified when we know the true value of the parameter, but this is rarely the case in patient studies. For this reason, animal, physical phantom and simulation studies have become an important tool in this optimization and evaluation process. However, it is desirable to validate these results using clinical studies and human images where the true value, which serves as a gold standard, is known imperfectly or not at all. Thus, in order to objectively compare these systems and algorithms, there is a need for evaluation methods that can be implemented in the absence of a gold standard.

We have developed a no-gold-standard approach to rank quantitative SPECT reconstruction methods when the end task is quantifying the mean activity concentration within a VOI [3]. The technique we have proposed extends on the basic framework that was initially proposed in Hoppin et al. [4] and further developed in Jha et al. [5]. The technique assumes a linear relationship between the estimated and true mean activity concentration values. It then used maximum-likelihood (ML) methods to estimate the parameters describing the linear relationship using the measured mean activity concentrations from the different reconstruction methods. We conducted several realistic simulation studies that demonstrated that the proposed no-gold-standard technique is able to successfully rank a set of reconstruction methods for quantitative SPECT [6].

An issue with the proposed technique is the requirement of a large number of patient studies to evaluate the reconstruction methods accurately. The validation studies that we conducted used the mean activity concentrations measured from up to 200 different VOIs, although we were able to reduce the number of patient studies requirement to 25 by using multiple organ VOIs from each patient. A similar issue has been observed with the previously developed no-gold-standard methods. For example, the evaluation of cardiac-ejection-fraction estimation algorithms using the no-gold-standard technique proposed in Hoppin et al. [4] used data from 85 patients [7]. Likewise, the simulation study for evaluating segmentation algorithms in DWMRI used images from up to 70 different simulated patient studies [5]. Often, such a large number of patient studies are not available. This requirement is a major obstacle in the wider clinical application of the no-gold-standard methods. A major endeavor of our research is to reduce this requirement.

A possible technique to reduce this requirement is by

incorporating prior information about the no-gold-standard parameters. In this regard, one could conduct physical phantom studies where the true value of the quantitative parameter is known. These experiments can yield information about a subset of the linear-relationship parameters. However, incorporating this information requires modifying the basic no-gold-standard method. In this paper, our objective was to develop methods to incorporate this prior information. We then studied whether incorporating this prior information led to improved performance, especially when the number of patient studies is small.

II. Theory

We consider the case where P patients are imaged using a SPECT imaging system. The patient images are reconstructed using K different reconstruction methods, and the mean activity concentration in the VOI is estimated. Our objective is to develop a no-gold-standard technique to evaluate these K reconstruction algorithms.

We assume that there is a linear relationship between the true and estimated activity values, consisting of both deterministic and random components. The linear relationship consists of a slope, intercept, and a zero-mean normally distributed noise term. We have confirmed this assumption of a linear relationship using realistic simulation studies for quantitative SPECT reconstruction methods with the aid of measures such as the Akiki information criterion (AIC) and Bayesian information criterion (BIC) [3]. Thus, denoting the true activity value for the p^{th} patient by a_p, the estimated activity value for this patient using the k^{th} reconstruction algorithm by $\hat{a}_p^k$, and the slope, intercept and standard deviation of the noise term by u_k,v_k,and σ_k, respectively, we write this model as

$$\hat{a}_p^k = u_k a_p + v_k + \mathcal{N}(0, \sigma_k), \tag{1}$$

where $\mathcal{N}(0, \sigma_k)$ denotes a zero-mean normally distributed random variable with standard deviation σ_k. We next assume that the true distribution has been sampled from some prior distribution. Based on empirical studies and in order to accommodate a wide variety of prior distributions, we choose this distribution to be a four-parameter beta distribution, parameterized by the vector $\mathbf{\Omega}$. Note that we do not know the values of $\{u_k, v_k, \sigma_k\}, k = 1, 2 \ldots K$ or $\mathbf{\Omega}$. For simplicity of notation, let us denote the set of all the observed data, i.e. the set $\{\hat{a}_p^k\}, k = \{1, 2 \ldots, K\}, p = \{1, 2 \ldots P\}$ by the vector $\hat{\boldsymbol{A}}$ and the set of linear model parameters $\{u_k, v_k, \sigma_k\}, k = 1, 2 \ldots K$ by the vector Θ. We can derive that

$$\text{pr}(\hat{\boldsymbol{A}}|\Theta, \mathbf{\Omega}\} = \int da_p \text{pr}(a_p|\Omega) \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left[\frac{-(a_p^k - a_p - v_k)^2}{2\sigma_k^2}\right]. \tag{2}$$

The original no-gold-standard method we previously proposed [3] uses a maximum-likelihood (ML) approach to estimate the values of those unknown parameters that maximize the probability of all the observed data, i.e.

$$\{\Theta, \mathbf{\Omega}\}_{\text{ML}} = argmax_{\Theta, \mathbf{\Omega}} \text{pr}(\hat{\boldsymbol{A}}|\Theta, \mathbf{\Omega}\}. \tag{3}$$

We observe that, in this procedure, we need to estimate a large number of parameters from a limited-sized data set. For example, if we are comparing three methods, we would have to estimate 13 parameters. Consequently, the estimated activity concentration $\hat{a}_p^k$ for many true values of a_p are required to perform this estimation task accurately. Further, to find a solution to this ML estimation problem, the optimization routine has to search over a large-dimensional space. There is a high degree of possibility that the solution could be trapped in a local extrema, thus yielding an inaccurate estimate.

One way to overcome these issues is to incorporate some prior information about the parameters Θ or $\mathbf{\Omega}$. This will help to regularize the ML estimation procedure. In this regard, we note that in a phantom-based experimental study, the true value of the underlying quantitative parameter is often known or can be measured using alternative techniques. Let us consider two such phantom-based experiments, where the true values of the mean activity concentration within a VOI were determined to be γ_1 and γ_2, respectively. Since these are phantom-based experiments, we could design the phantom to have to have a large value of mean activity concentration in the VOI and acquire the data long enough, so that we obtain a large number of photon counts. In this case, the measurement noise can be much reduced in comparison to patient studies. Let us denote the estimated values for the mean activity concentrations γ_1 and γ_2 after using the k^{th} reconstruction method by $\hat{\gamma}_1^k$ and $\hat{\gamma}_2^k$, respectively. Then, using the assumption of linearity as in Eq. (1), we can write

$$\hat{\gamma}_i^k = u_k \gamma_i + v_k + \mathcal{N}(0, \varsigma), \ i = 1, 2 \tag{4}$$

where ς is the standard deviation of the reduced measurement noise in these experiments. From the estimates $\hat{\gamma}_i^k$ and using Eq. (4) an estimate of u_k, denoted by $\hat{u}_k$ is given by

$$\hat{u}_k = \frac{\hat{\gamma}_2^k - \hat{\gamma}_1^k}{\gamma_1 - \gamma_2}. \tag{5}$$

Since $\hat{u}_k$ is the difference of two normally distributed random variables, we can derive that $\hat{u}_k$ will be normally distributed with mean u_k and standard deviation $\sqrt{2}\varsigma/|\gamma_1 - \gamma_2|$ [8]. Similarly, we can compute an estimate of v_k, denoted by $\hat{v}_k$, using Eq. (4), as follows:

$$\hat{v}_k = \frac{\hat{\gamma}_2^k \gamma_1 - \hat{\gamma}_1^k \gamma_2}{\gamma_1 - \gamma_2}. \tag{6}$$

The random variable $\hat{v}_k$ is again the difference of two normally distributed random variables and thus will also be normally distributed with mean v_k and standard deviation $\frac{\varsigma}{|\gamma_1 - \gamma_2|}\sqrt{\gamma_2^2 + \gamma_1^2}$.

We repeat these phantom studies N times with the true value of activity concentration as γ_1 and γ_2 so that we can measure ς precisely to the extent that we can assume it as a known value. Further, we compute the sample means of $\hat{u}_k$ and $\hat{v}_k$ from the multiple phantom experiments, which we denote by $\bar{\hat{u}}_k$ and $\bar{\hat{v}}_k$, respectively. In that case, it can be shown that when we do not have any prior information about u_k, the posterior

distributions of u_k and v_k are also normally distributed such that [9]

$$\mathrm{pr}(u_k|\hat{u}_k,\varsigma) = \mathcal{N}\left(\bar{\hat{u}}_k, \sqrt{\frac{2}{N}}\frac{\varsigma}{|\gamma_1-\gamma_2|}\right). \quad (7a)$$

$$\mathrm{pr}(v_k|\hat{v}_k,\varsigma) = \mathcal{N}\left(\bar{\hat{v}}_k, \frac{\varsigma}{N|\gamma_1-\gamma_2|}\sqrt{\gamma_2^2+\gamma_1^2}\right). \quad (7b)$$

Thus, from these phantom-based experiments, we can obtain a prior probability distribution for the variables u_k and v_k.

To incorporate this prior information, we estimate the maximum-a-posteriori (MAP) solution of the no-gold-standard parameters. This estimate is given by

$$\{\Theta,\Omega\}_{\mathrm{MAP}} = \mathrm{argmax}_{\Theta,\Omega}\ \mathrm{pr}(\Theta,\Omega|\hat{\boldsymbol{A}}). \quad (8)$$

Applying Bayes theorem yields

$$\mathrm{pr}(\Theta,\Omega|\hat{\boldsymbol{A}}) = \frac{\mathrm{pr}(\hat{\boldsymbol{A}}|\Theta,\Omega)\mathrm{pr}(\Theta,\Omega)}{\mathrm{pr}(\hat{\boldsymbol{A}})}. \quad (9)$$

Substituting this expression in Eq. (8) and recognizing that the denominator in the above equation is just a normalization term, we obtain

$$\{\Theta,\Omega\}_{\mathrm{MAP}} = \mathrm{argmax}_{\Theta,\Omega}\ \mathrm{pr}(\hat{\boldsymbol{A}}|\Theta,\Omega)\mathrm{pr}(\Theta,\Omega). \quad (10)$$

The prior information obtained from the phantom experiments is accounted for by recognizing the independence of the no-gold-standard parameters and thus rewriting the expression for the prior distribution $\mathrm{pr}(\Theta,\Omega)$ as

$$\mathrm{pr}(\Theta,\Omega) = \prod_{k=1}^{K}\mathrm{pr}(u_k)\mathrm{pr}(v_k)\mathrm{pr}(\sigma_k)\mathrm{pr}(\Omega). \quad (11)$$

The expressions for $\mathrm{pr}(u_k)$ and $\mathrm{pr}(v_k)$ are obtained from the phantom study (Eq. (7)). Also, the expression for the likelihood of the data, i.e. $\mathrm{pr}(\hat{\boldsymbol{A}}|\Theta,\Omega)$ is obtained from Eq. (2). We thus obtain a MAP estimate of the no-gold-standard parameters. We denote the estimates of the slope, intercept and standard deviation of noise terms by $\hat{u}_{\mathrm{MAP},k}$, $\hat{v}_{\mathrm{MAP},k}$ and $\hat{\sigma}_{\mathrm{MAP},k}$, respectively.

We use the parameters obtained from the no-gold-standard technique to compute two figures of merit: Noise-to-slope ratio $\hat{\sigma}_{\mathrm{MAP},k}/\hat{a}_{\mathrm{MAP},k}$, which represents the precision of the algorithm when the slope and intercepts are treated as calibration factors and corrected for, and the scaled ensemble mean square error ($\mathrm{EMSE}_{\mathrm{sc}}$), which is defined as

$$\mathrm{EMSE}_{\mathrm{sc}} = \frac{\hat{v}^2_{\mathrm{MAP},k}+\hat{\sigma}^2_{\mathrm{MAP},k}}{\hat{u}^2_{\mathrm{MAP},k}}. \quad (12)$$

$\mathrm{EMSE}_{\mathrm{sc}}$ represents the ensemble mean square error (EMSE) with the algorithm when the slope obtained from the no-gold-standard method is treated as a calibration factor and used to correct the data. This metric measures both the bias and variance of the reconstruction algorithm.

III. Experiments and Results

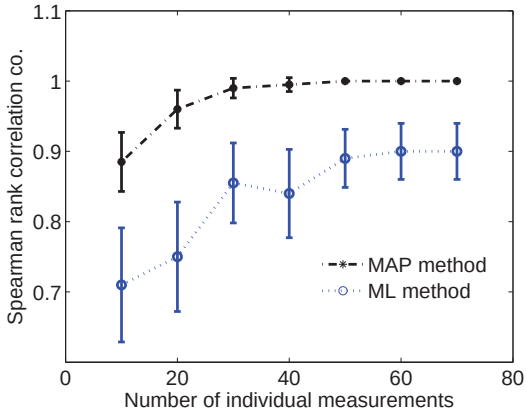
Simulation studies were conducted to study the performance of the proposed no-gold-standard technique and compare it with the ML-based no-gold-standard technique. We generated the true mean activity concentration value within the VOI for a certain number of patient studies by sampling a beta distribution. The estimated mean activity concentrations for three assumed reconstruction methods were obtained using Eq. (1) with different values for the slope, intercept, and the noise standard deviation parameters. We refer to the mean activity concentration within a particular VOI estimated using a particular reconstruction method as an individual measurement from a clinical study, or more simply, an individual measurement. These estimated values were input to the ML and MAP-based no-gold-standard techniques. From the estimated figures of merit, the rankings of the three methods were obtained. The correlation of the obtained and true rankings of the three methods was quantified by computing the Spearman's rank correlation coefficient between the true and estimated rankings. This process was repeated for 100 noise realizations, where each noise realization corresponded to obtaining the mean activity concentration estimates using the three reconstruction methods for the same set of underlying true mean activity concentration values. This experiment was repeated for different numbers of patient studies, or equivalently, with different number of individual measurements.

The correlation coefficients of the rankings with the two figures of merit as a function of the number of true values are shown in Fig. 1a and 1b. We first observed that the proposed MAP-based technique was able to successfully rank the three methods accurately for even smaller numbers of individual measurements. For example, in the experiments with the MAP-based method, just 40 individual measurements yielded rankings with the mean Spearman's ranking correlation coefficient equal to 0.99 ± 0.05 and 0.99 ± 0.07, using noise-to-slope ratio and $\mathrm{EMSE}_{\mathrm{sc}}$ as the figures of merit, respectively. This implies that the proposed technique did not correctly predict the rankings for only one out of 100 noise realizations. This result is noteworthy since multiple individual measurements of mean activity concentrations can be obtained from the same patient by considering multiple organ VOIs. For example, in our previous study, we obtained 8 individual measurements from each patient study [6]. Thus, this result indicates that the number of patient studies requirement could be reduced significantly.

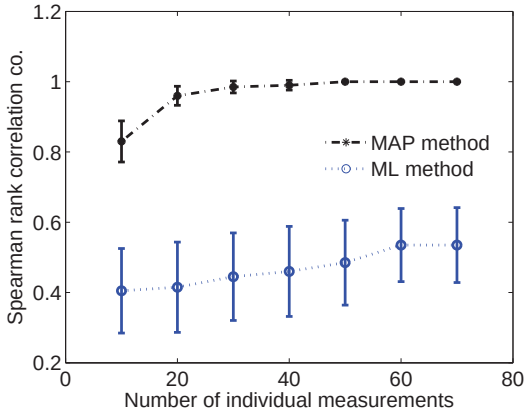
Additionally, the MAP-based method clearly performed better than the ML method when the number of individual measurements, i.e., the number of patient studies, was small. As the number of patient studies increased, the performance of the MAP and ML-based method improved, but the MAP-method consistently outperformed the ML-based method.

IV. Conclusions

We have developed a MAP-based no-gold-standard method to evaluate reconstruction methods for quantitative SPECT imaging when the task is estimating the activity concentration



(a)



(b)

Fig. 1: The rank correlation coefficient between the true and estimated rankings plotted as a function of the number of individual measurements for the ML and MAP-based no-gold-standard techniques. The error bars denote the 95% confidence intervals. The rankings were determined using (a) The noise-to-slope ratio and (b) $\mathrm{EMSE}_{\mathrm{sc}}$ parameters.

in a VOI. The method incorporates prior information about the estimated no-gold-standard parameter values. We also proposed a technique to obtain this prior information from phantom-based experimental studies. Preliminary results indicate that, for small numbers of patient studies, the proposed MAP-based method ranked the methods more accurately than the ML-based no-gold-standard technique. We are currently studying the efficacy of this technique for assessing specific quantitative SPECT reconstruction methods. By reducing the required number of patient studies, the proposed MAP-based no-gold-standard method can enable wider clinical application of objective evaluation in the absence of ground truth with clinical data. The proposed MAP-based framework can also be used to incorporate any other available prior information about the no-gold-standard parameters. Further, the method can be extended to objectively evaluate quantitative SPECT systems and algorithms for other quantitative tasks such as VOI segmentation and histogram estimation.

V. Acknowledgment

This work was supported by National Institute of Biomedical Imaging and Bioengineering of National Institute of Health under numbers R01-EB016231 and EB013558.

References

[1] D. L. Bailey and K. P. Willowson, "An evidence-based review of quantitative SPECT imaging and potential clinical applications," *J. Nucl. Med.*, vol. 54, no. 1, pp. 83–89, Jan 2013.

[2] ——, "Quantitative SPECT/CT: SPECT joins PET as a quantitative imaging modality," *Eur. J. Nucl. Med. Mol. Imaging*, Sep 2013.

[3] A. K. Jha, N. Song, and E. Frey, "No-gold-standard methods for objective evaluation of reconstruction methods for quantitative spect imaging," in *IEEE Medical Imaging conference*, 2014.

[4] J. W. Hoppin, M. A. Kupinski, G. A. Kastis, E. Clarkson, and H. H. Barrett, "Objective comparison of quantitative imaging modalities without the use of a gold standard," vol. 21, pp. 441–9, May 2002.

[5] A. K. Jha, M. A. Kupinski, J. J. Rodriguez, R. M. Stephen, and A. T. Stopeck, "Task-based evaluation of segmentation algorithms for diffusion-weighted mri without using a gold standard," *Phys Med Biol*, vol. 57, no. 13, pp. 4425–4446, Jul 2012.

[6] A. K. Jha and E. Frey, "Validating a no-gold-standard technique to assess quantitative spect methods," 2015.

[7] M. A. Kupinski, J. W. Hoppin, J. Krasnow, S. Dahlberg, J. A. Leppo, M. A. King, E. Clarkson, and H. H. Barrett, "Comparing cardiac ejection fraction estimation algorithms without a gold standard," *Acad. Radiol.*, vol. 13, pp. 329–37, Mar 2006.

[8] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. Tata McGraw-Hill, 2002.

[9] P. Diaconis and D. Ylvisaker, "Conjugate priors for exponential families," *Annals Stat.*, vol. 7, no. 2, pp. 269–81, 1979.