# Predicting nucleus basalis of Meynert volume from compartmental brain segmentations

Hennadii Madan[1], Rok Berlot[2], Nicola J Ray[3], Franjo Pernuš[1], and Žiga Špiclin[1]

[1]Faculty of Electrical Engineering, Laboratory of Imaging Technologies,
University of Ljubljana, Tržaška 25, 1000 Ljubljana, Slovenia
[2]Department of Neurology, University Medical Centre Ljubljana,
Zaloška 2, 1000 Ljubljana, Slovenia
[3]Department of Psychology, Manchester Metropolitan University,
53 Bonsall Street, Manchester, M15 6GX, UK
{hennadii.madan,franjo.pernus,ziga.spiclin}@fe.uni-lj.si

**Abstract.** In clinical practice one often encounters a situation when a quantity of interest cannot be measured routinely, for reasons such as invasiveness, high costs, the need for special equipment, etc. For instance, research showed that early cognitive decline can be predicted from volume (atrophy) of the nucleus basalis of Meynert (NBM), however its small size makes it difficult to measure from brain magnetic resonance (MR) scans. We treat NBM volume as an unobservable quantity in a statistical model, exploiting the structural integrity of the brain, and aim to estimate it indirectly based on one or more interdependent, but possibly more accurate and reliable compartmental brain volume measurements that are easily accessible. We propose a Bayesian approach based on the previously published reference-free error estimation framework to achieve this aim. The main contribution is a novel prior distribution parametrization encoding the scale of the distribution of the unobservable quantity. The proposed prior is more general and better interpretable than the original. In addition to unobservable quantity estimates, for each observable we calculate a figure of merit as an individual predictor of the unobservable quantity. The framework was successfully validated on synthetic data and on a clinical dataset, predicting the NBM volume from compartmental segmentations of structural brain MR images, based on volumes of the whole-brain and hippocampal subfields.

**Keywords:** Bayesian Inference, Markov Chain Monte Carlo, Validation, Brain Segmentation, Clinical Dataset, Magnetic Resonance Imaging.

## 1 Introduction

Computational analysis of medical images is increasingly used for extracting quantitative imaging biomarkers (QIBs) — scalar measurements that characterize a certain morphological or functional aspect of the anatomy of interest. In certain diseases there exist QIBs that allow for disease diagnosis at an early stage, well before clinical symptoms appear. For instance, recent research showed

that early cognitive decline can be predicted by measuring volume (atrophy) of nucleus basalis of Meynert (NBM) [4], however its small size makes it a difficult target for computational or even manual segmentation on resolution–limited magnetic resonance (MR) brain images.

Based on the integrity of the brain, one may treat the difficult to measure NBM as an unobservable quantity and model it in terms of one or several interdependent observed routine compartmental brain volume measurements. Two questions are then of interest: (a) which observed quantities are the best "predictors" of the unobservable quantity and (b) what are the likely values of the unobservable quantity given the observed ones?

To answer these questions we propose a novel Bayesian approach based on the reference-free error estimation framework [3]. The original framework was designed to compare measurement methods (MMs) for the same quantity and has assumptions specific to MMs that are necessary for model identification (see section 2.1). We lift these assumptions by using additional information about the distribution of the unobservable quantity (section 2.2) and use the reference-free estimates to answer the questions (a) and (b) for a synthetic and a clinical dataset (section 3).

## 2 Reference-free error estimation

Let $q$ denote the unobservable quantity in patient $p$, $p = 1..N$. Assume that we have $M$ easy-to-measure observables, indexed with $m$, $m = 1..M$. Let $y_{pm}$ be the value of $m$-th observable in patient $p$ defined by the value of a certain deterministic function $g_m(q_p)$, corrupted by random noise $\epsilon$:

$$y_{pm} = g_m(q_p) + \epsilon_{pm} \tag{1}$$

Assuming that $g_m$ are analytic and we are dealing with values of $q_p$ from a finite interval $[\underline{q_p}, \overline{q_p}]$, we may approximate $g_m$ with a $K - th$ degree polynomial representing truncated Taylor series about a point $q_o \in [\underline{q_p}, \overline{q_p}]$:

$$y_{pm} = \sum_{k=0}^{K} b_{km}(q_p - q_0)^k + \epsilon_{pm} = \sum_{k=0}^{K} b_{km} x_p^k + \epsilon_{pm} \tag{2}$$

where notation $x_p \triangleq q_p - q_0$ is introduced for brevity. Multivariate Gaussian (MVG) distribution is assumed for random errors:

$$\boldsymbol{\epsilon_p} \sim \mathcal{N}(0, \Sigma) \tag{3}$$

where $\boldsymbol{\epsilon_p} = (\epsilon_{p1}, ..., \epsilon_{pM})^\top$ and $\Sigma$ is a covariance matrix.

From (2) and (3) the likelihood of observing $\boldsymbol{y_p} \triangleq [y_{p1}, ..., y_{pM}]$ is:

$$l_p \triangleq f(\boldsymbol{y_p} \mid B, \Sigma, x_p) = \mathcal{N}(B\boldsymbol{\chi}, \Sigma) \tag{4}$$

where $f$ denotes probability density, $B \triangleq [b_{km}] \in \mathbb{R}^{KM}$ and $\boldsymbol{\chi} \triangleq [1, x_p, x_p^2, \ldots, x_p^K]$. The likelihood for the entire set of observations is then:

$$l \triangleq f(Y \mid \theta) = \prod_{p=1}^{N} l_p \tag{5}$$

where $Y = [y_{pm}] \in \mathbb{R}^{N \times M}$, $\theta = \{B, \Sigma, \boldsymbol{x}\}$, $\boldsymbol{x} = [x_1, ..., x_N]$. By Bayes' Theorem the posterior probability density of $\theta$ given the observed values $Y$ is:

$$f(\theta \mid Y) \propto l \cdot f(\theta) \tag{6}$$

where $f(\theta)$ is the prior probability density of model parameters. When both $l$ and $f(\theta)$ are specified, one can draw samples from $f(\theta \mid Y)$ using Markov chain Monte-Carlo (MCMC). The samples allow estimation of marginal posterior expectations of quantities of interest with associated uncertainties. For instance the quantity

$$F_m \triangleq \max_{q \in [\underline{q_p}, \overline{q_p}]} \frac{\mid g'_m(q) \mid}{\sigma_m} \tag{7}$$

can be shown to be equal to the reciprocal of the smallest possible (over $q \in [\underline{q_p}, \overline{q_p}]$) root mean square error (RMSE) of $q_p$ estimates, obtained from $y_{pm}$ for a particular $m$ by inversion of $g_m$. In other words, it is the reciprocal value of the smallest error one would make if one used only the $m$-th observable (with known $g_m$) to estimate $q$. Quantity in (7) can be interpreted as a figure of merit of the $m$-th observable as a predictor of $q$, thereby providing an answer to the question (a), while the estimates of $q_p$ provide an answer to the question (b), as posed in the Introduction.

## 2.1 Priors in previous works

The likelihood $l$ can be shown to be degenerate. This means that, in order to identify the model, the priors must be sufficiently informative. In previous research, uniform priors on $q_p$ were used in conjunction with peaked informative priors on $b_{km}$:

$$q_p \sim \mathcal{U}(\underline{q_p}, \overline{q_p}), \quad \forall p, \tag{8}$$

$$b_{1m} \sim \mathcal{N}(1, \sigma_{b1}), \quad \forall m, \tag{9}$$

$$b_{km} \sim \mathcal{N}(0, \sigma_{bk}), \quad \forall m, k \neq 1, \tag{10}$$

where $\underline{q_p}$ and $\overline{q_p}$ are either physical or physiological bounds on $q_p$ and $\sigma_{bk} = c_k(\overline{q_p} - \underline{q_p})^{-k}$, $c_k \approx 1$. Such priors are based on the assumption that $y_{mp}$ are (imperfect) measurements of $q_p$ and therefore $g_m$ are close to identity at least in the vicinity of $q_0$. To parametrize the covariance matrix a noise variance-correlation separation strategy [1] based on the following decomposition was applied:

$$\Sigma = SRS \tag{11}$$

where $S = diag([\sigma_1, ..., \sigma_M])$ is a diagonal matrix of random error standard deviations (STDs) and $R = [r_{ij}]$ a symmetric correlation matrix. Then, STDs were assigned truncated Jeffreys priors:

$$f(\sigma_m) = \begin{cases} \frac{1}{\sigma_m}, & \underline{\sigma_m} < \sigma_m < \overline{\sigma_m} \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Truncation guaranteed that the posterior was proper, $\underline{\sigma_m}$ was set to measurement resolution, while $\overline{\sigma_m}$ was set to $(\max_p y_{pm} - \min_p y_{pm})$. The correlation matrix was assigned LKJ priors [2] with $\eta = 1$.

This approach was successfully validated on a clinical in vivo dataset and several synthetic datasets [3]. Despite the usefulness of informative priors on $b_{km}$ in the context of MM comparison, these priors are not applicable for general $g_m(q)$.

## 2.2  Proposed prior

The main contribution in this paper is a novel encoding of prior distribution that lifts assumptions on $b_{km}$ that constrain $g_m$ to be close to identity. Instead, the novel prior draws on additional information about the scale of $q_p$ distribution to identify the model. The additional information required is (i) an interval of likely values of the minimum and maximum (over $p$) points in the $x_p$ sample encoded as the following conditions:

$$\max_p q_p \geq \overline{q_p} - \overline{\epsilon}$$
$$\min_p q_p \leq \underline{q_p} + \underline{\epsilon} \tag{13}$$

where $\overline{\epsilon}, \underline{\epsilon} > 0$ are a priori limits on how far the minimum and the maximum values of $q_p$ might reside from the boundaries of the specified uniform prior on $q_p$, defining the *magnitude* of the scale of $q_p$ distribution; and (ii) a pair of indices $\overline{p}$ and $\underline{p}$ for which it is known that $q_{\overline{p}} > q_{\underline{p}}$, disambiguating the *sign* of the scale of $q_p$ distribution and determining the order of $q_p$ estimates w.r.t. their true values. The points $\overline{p}$ and $\underline{p}$ need not coincide with the maximum and minimum points. Polynomial coefficients $b_{km}$ are assigned flat priors, while the priors on $\Sigma$ are left unchanged as per (11) and (12).

## 3  Validation

The capability of the proposed framework to estimate the values of an unobservable quantity based on several related and interdependent quantities was validated on datasets of synthetic and clinical scalar measurements.

We focus on the ability of the framework to estimate $F_m$ in (7) and $q_p$. Each plot of $F_m$ also reports correlation coefficient $\rho$ with respect to reference values:

$$\rho \triangleq \frac{\sum_{m=1}^{M}(\tilde{F}_m - \langle \tilde{F}_m \rangle)(F_m^* - \langle F_m^* \rangle)}{\sum_{m=1}^{M}(\tilde{F}_m - \langle \tilde{F}_m \rangle)^2 (F_m^* - \langle F_m^* \rangle)^2}, \tag{14}$$

where $\tilde{F}_m$ is the posterior estimate and $F_m^*$ is the value obtained by least squares regression against *known* reference values of $q$, used in validation. Plots for $q_p$ also provide the RMSE of the estimates:

$$A \triangleq \sqrt{\sum_{p=1}^{n} q_p^2 / N} \tag{15}$$

and the smallest RMSE one would obtain if he or she used only the best predictor with known polynomial coefficients:

$$A_1 \triangleq \min_m 1/F_m^*. \tag{16}$$

### 3.1 Synthetic data

Experiment with synthetic data was conducted to demonstrate the ability of the framework to estimate the model parameters with highly non-linear $g_m$ that would otherwise have invalidated the assumptions of the original reference-free error estimation scheme [3]. We have generated $N = 30$ points from $\mathcal{U}(0, 55)$ and, at those points, evaluated polynomials with coefficients given in table 1. The obtained values were then perturbed with MVG noise with standard deviations $\sigma_m$ and correlation matrix $R$ from table 1.

Table 1: Parameters used to generate synthetic data.

| $m$ | $b_{0m}$ | $b_{1m}$ | $b_{2m}$ | $\sigma_m$ |
|---|---|---|---|---|
| 1 | 80 | $-4.0$ | 0.01 | 7.0 |
| 2 | $-80$ | 4.0 | $-0.02$ | 6.2 |
| 3 | $-80$ | 12.0 | $-0.20$ | 4.3 |
| 4 | 40 | $-3.5$ | 0.04 | 2.2 |

$$R = \begin{pmatrix} 1 & 0.9 & 0 & 0 \\ 0.9 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The parameters of the prior were setup as follows: $\underline{q_p} = 0$, $\overline{q_p} = 55$, $\underline{\epsilon} = \overline{\epsilon} = 5$, $\underline{\sigma_m} = 0.001$, $\overline{\sigma_m} = 55$, indices $\overline{p}$ and $\underline{p}$ were picked at random. $K$ was set to 2, $q_0$ was set to 0.

Results are given in fig. 1. Figure of merit estimates are in agreement with the true values and as such allow to answer the question (a). The unobservable quantity estimates are very close to the known true values and can thus be used to answer the question (b). Although $A > A_1$, it must be understood that $A$ is actual RMSE calculated over $[\underline{q_p}, \overline{q_p}]$ , while $A_1$ assumes that $q$ is in the optimal region for the particular predictor that produces the smallest value of this quantity.
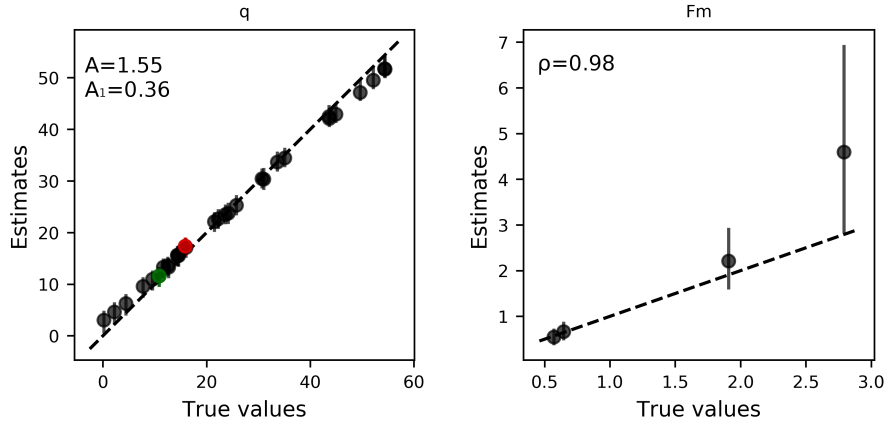
Fig. 1: Reference-free estimates of the unobservable variable $q$ and the figure of merit $F_m$ compared to corresponding known true values for the synthetic dataset. *Red and green points* were used to disambiguate the slope sign in regression model (see section 2.2). See section 3 for definitions of $A$, $A_1$ and $\rho$. *Dashed lines* indicate optimal matching of estimated and true values.

## 3.2 Clinical data

Structural T1-weighted MR scans of a group of 40 patients, including 20 healthy elderly and 20 with mild cognitive impairment, the prodromal stage of Alzheimer's disease were analysed. Data for analysis consisted of volumetric measurements of whole-brain, hippocampus and its subfields, obtained using Freesurfer and DARTEL segmentation tools. These data were used to attempt to predict the NBM volume.

The NBM is a small region that is not routinely measured, yet it is associated with cognitive health and implicated in various neurodegenerative disorders. For validation purposes, the NBM volume was extracted using a detailed stereotactic atlas. All volumetric measurements were normalized to the total intracranial volume to account for the differences in head size between subjects.

The minimum and the maximum points were determined from normalized reference NBM volumes and provided indices $\underline{p}$ and $\overline{p}$, based on which the remaining parameters of the prior were setup: $\underline{\epsilon}$ and $\overline{\epsilon}$ were set to 0.02, while $\underline{q_p}$ and $\overline{q_p}$ were set so that the respective minimum and maximum values were approximately at $\underline{q_p} + \underline{\epsilon}/2$ and $\overline{q_p} + \overline{\epsilon}/2$. $K$ was set to 1, $q_0$ was set to $(\underline{q_p} + \overline{q_p})/2$.

The resulting estimates in fig. 2 show good agreement with the reference. The RMSE was slightly lower as compared to the RMSE obtainable from a single predictor, thus successfully answering the question (b). Generally, estimates of $F_m$ were in good agreement with those obtained using least squares on reference values, taking into account the associated uncertainty, and therefore enable one to answer question (a).
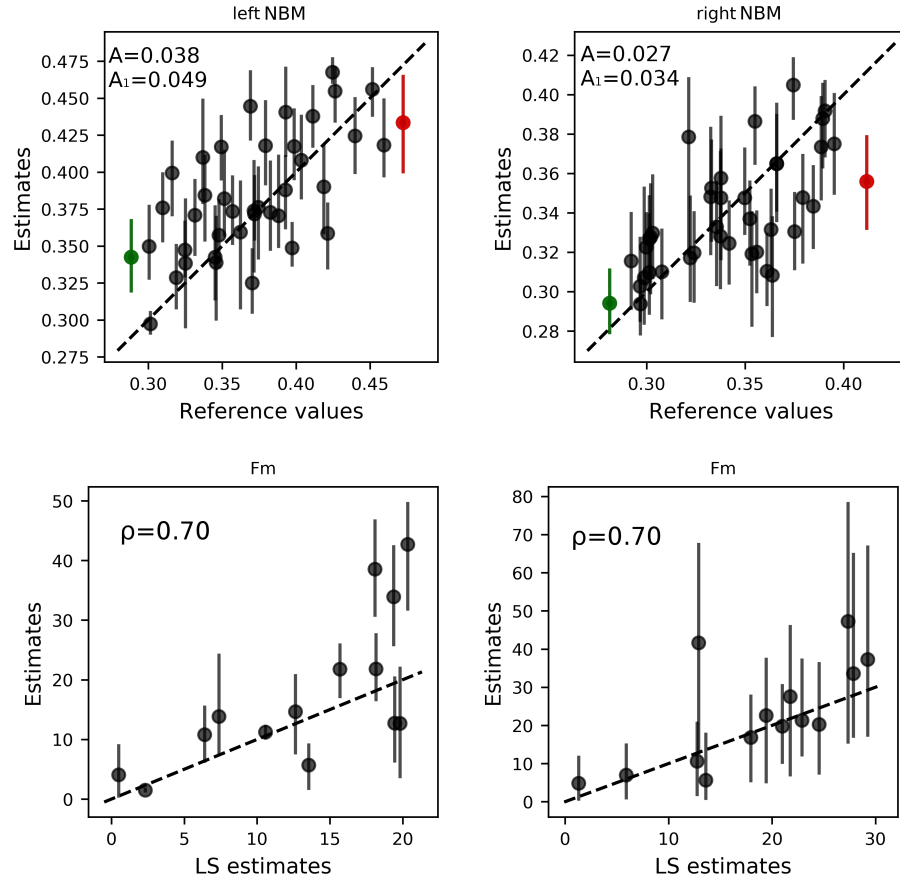
Fig. 2: *Top:* estimates of normalized NBM volume from volumes of hippocampal sub-fields for each hemisphere plotted against reference values. *Green and red points* denote $\underline{p}$ and $\overline{p}$ (section 2.2). *Bottom:* figure of merit $F_m$ estimates of hippocampal subfields as predictors for NBM volume plotted against the estimates obtained by least squares regression with respect to reference values. See section 3 for definitions of $A$, $A_1$ and $\rho$. *Dashed lines* are identity lines indicating perfect match.

## 4    Discussion

Results show that the proposed priors for the reference-free error estimation framework produce valid estimates of the unobservable quantity (NBM) and identify best easy-to-measure predictors of this quantity. The proposed priors are more general, but at the same time are more practical and objective than the original ones [3]. Instead of vague guesses about coefficients of Taylor expansion one has to provide interpretable, clearly defined parameters: span of the unobservable quantity, its uncertainty in the form of tolerance parameters and a pair of point indices with known value ordering.

For many biomarkers these parameters can be measured or inferred. For example, lower bound of the biomarker distribution may defined exactly by including a healthy control subject in the dataset. If the patient with the highest value of the biomarker can be identified, then by a single application of a (possibly expensive) gold standard MM, the uncertainty of the upper bound can be reduced to the level of the method's nominal accuracy. To specify the pair of order-defining indices one may again use controls: for a large class of biomarkers a healthy control subject will have the value of the biomarker equal to zero, which is guaranteed to be less than that of a patient who has the relevant medical condition. Another possibility is to use up to two applications of a gold standard method.

Whenever applicable, reference-free error estimation provides significant savings of time and costs normally associated with reference measurements: those of human operators, non-standard protocols, high-end acquisition equipment, material costs (e.g. contrast agents, materials of phantoms), instrumentation (frames, fiducial markers), administrative overhead, patient recovery and side effects from invasive measurements, making it an invaluable analytical tool.

## References

1. Barnard, J., McCulloch, R., Meng, X.L.: Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. Statistica Sinica pp. 1281–1311 (2000)
2. Lewandowski, D., Kurowicka, D., Joe, H.: Generating random correlation matrices based on vines and extended onion method. Journal of multivariate analysis 100(9), 1989–2001 (2009)
3. Madan, H., Pernuš, F., Žiga Špiclin: Reference-free error estimation for multiple measurement methods. Statistical Methods in Medical Research 0(0), 0962280217754231 (2018), https://doi.org/10.1177/0962280217754231, pMID: 29384043
4. Ray, N.J., Bradburn, S., Murgatroyd, C., Toseeb, U., Mir, P., Kountouriotis, G.K., Teipel, S.J., Grothe, M.J.: In vivo cholinergic basal forebrain atrophy predicts cognitive decline in de novo Parkinson's disease. Brain 141(1), 165–176 (Jan 2018)