Predicting Multi-coil RF Shims via Machine Learning

Julianna D. Ianni[1,2], Zhipeng Cao[1,2], and William A. Grissom[1,2,3,4]
[1]Vanderbilt University Institute of Imaging Science, [2]Department of Biomedical Engineering, [3]Department of Radiology,
[4]Department of Electrical Engineering, Vanderbilt University, Nashville Tennessee, 37232

INTRODUCTION

At high MRI field strengths (7T+), the transmit RF field (B_1^+) becomes very non-uniform, and resulting images contain unwanted shading. Patient-tailored radiofrequency (RF) shimming is capable of correcting for this inhomogeneity on a patient-by-patient basis by adjusting the magnitudes and phases of each coil in a multicoil array.[1,2] Higher numbers of coils provide more degrees of freedom to shape the transmit field.[3]

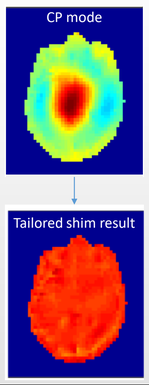


Fig 1. Example of inhomogeneous transmit field in brain at 7T (simulated). Circularly polarized (CP) mode is shown versus the tailored result.

<u>The *current* workflow increases time in the scanner.</u>
1. Measure patient B_1^+ maps: *This is scan-time intensive especially with large-numbers of coils.*
2. Calculate optimal RF weights: *Solving the (non-convex) magnitude least squares (MLS) problem is a computationally intensive step, and it is difficult to find a global optimum.*[4]

<u>The *proposed* workflow mitigates this:</u>
1. Measure patient B_1^+ maps (or subset of typical B_1^+ map data); extract B_1^+ features
2. Apply a learned transform to <u>predict RF shims directly</u> – *computationally cheap*

METHODS

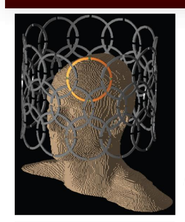


Fig. 2. B_1^+ maps were simulated for a 36ch coil at 7T using XFDTD.

<u>Head phantoms</u> B_1^+ maps were simulated for phantom head morphologies (1 male and 1 female) derived from the Virtual Family[5] for a 36-channel coil at 7T using XFDTD (Remcom Inc., State College, PA, USA) (Fig. 2). Ten simulations were performed, *with different magnifications of the original 2 phantoms*. Each phantom's mask in 3 dimensions is shown in Figure 5 (right) along with the magnifications applied to the original phantom dimensions. This resulted in a total of 310 axial slices of B_1^+ data with 2mm isotropic resolution.

<u>Shim design</u>: A variable exchange magnitude least squares (MLS) method[4,6] with 100 random starts was used to design shims for each axial slice. Solutions with lowest B_1^+ inhomogeneity were used to train and test several learning methods using 10-fold cross-validation.

<u>Training Feature Set</u>

For each slice, features used for training consisted of:
1) The standard deviations of the within-mask x & y coordinates
2) The centroid of the brain mask in x & y
3) The central Fourier Transform coefficients of the B_1^+ field maps for each coil
4) The z-position of the slice within the coil
5) 1st order cross-terms of the above features.
6) The sign of (1-4) above.

Features were normalized to zero mean and unit standard deviation.

The effect of varying the number of Fourier Transform coefficients used for training was explored, and two feature set configurations were compared:
1. Case 1: Neighborhood of 3 coefficients (Fig 3), no cross-terms.
2. Case 2: Neighborhood of 1 coefficient, no sign features

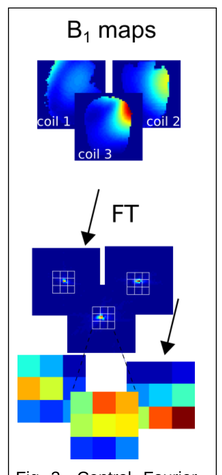
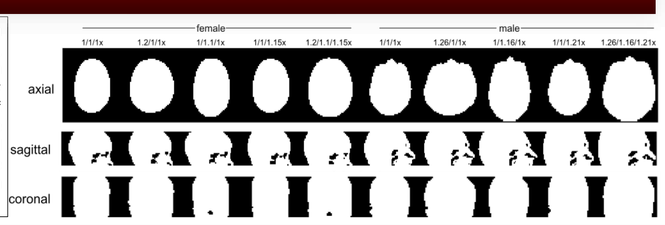


Fig 3. Central Fourier-transform coefficients are extracted from B_1^+ maps for each coil, slice, and subject.

HEAD PHANTOMS

Figure 4. Masks for each phantom in axial, sagittal , and coronal planes. Each phantom was generated by a magnification in x/y/z of either the male or female phantom from the Virtual Family. Magnifications are shown above each phantom.



LEARNING METHODS

Learning methods

Direct learning methods
- Nearest neighbors (NN)
- k-nearest neighbors (kNN)
  -with 5 neighbors
- Kernelized nearest neighbors (KNN)
  -Using a Gaussian kernel with unit standard deviation
- Principal components regression (PCR)
  -with 10 components
  -1 run split into real/imaginary components
  -1 run split into magnitude/phase (PCRmp)

POCS shimming methods
- POCS Ridge regression (POCS-RR)
- POCS Random Forest (POCS-RF)

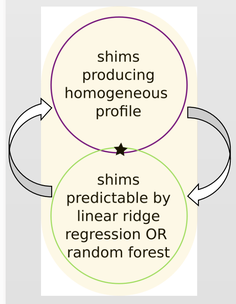


Figure 5. POCS shim design process. Algorithm iterates over 2 projections: shims minimizing profile error and those predictable by (a) linear ridge regression or (b) random forest regression of the features. The phase profile is updated after projection onto the 2nd set.

Obtaining more trainable solutions: projection over convex sets (POCS) shimming

-Shims are *<u>designed to be predictable</u>* by ridge regression of features, using a POCS iterating over several projections:

- <u>Projection 1.</u> A few conjugate-gradient (CG) steps are performed to update shims using the variable-exchange MLS method.
- <u>Projection 2.</u> Shims are projected onto the set predictable by
A) Linear ridge regression (POCS-RR) or
B) Random forest regression[7] (POCS-RF)

and the new target phase is set to that of the profile generated by the current shim solution.

- Termination: is determined in the 1st fold of a 10-fold cross-validation; POCS iterations are stopped when the shim profile error begins increasing (to avoid overfitting), and all remaining folds stop at the same number of iterations.

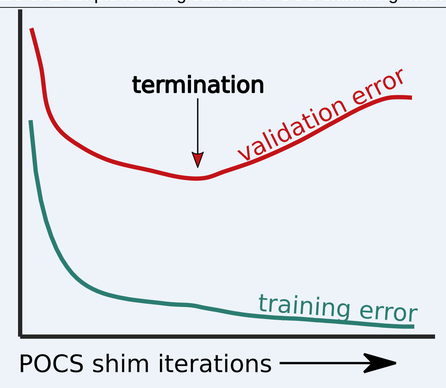
Predictions are made by applying the feature projection learned from (and enforced in) the training shims.

...But not too trainable:
Mitigating overfitting with POCS:

The POCS shimming methods are capable of overfitting the training data. Steps are taken to avoid this:

1. Run POCS shimming with 1st fold of 10-fold cross-validation, calculating validation error (for the data not trained on) concurrently:

Figure 6. Example learning curve for POCS shimming methods.



2. Stop when validation error increases, note the number of iterations = X & *exclude this validation data from all test sets*.
3. Run POCS shimming for remaining 9 folds, stopping at X iterations blindly.

RESULTS

Excitation Profiles

a) **Case 1:** B_1^+ neighborhood =3, + sign features, POCS run to convergence

b) **Case 2:** B_1^+ neighborhood =1, + cross-term features, POCS early termination

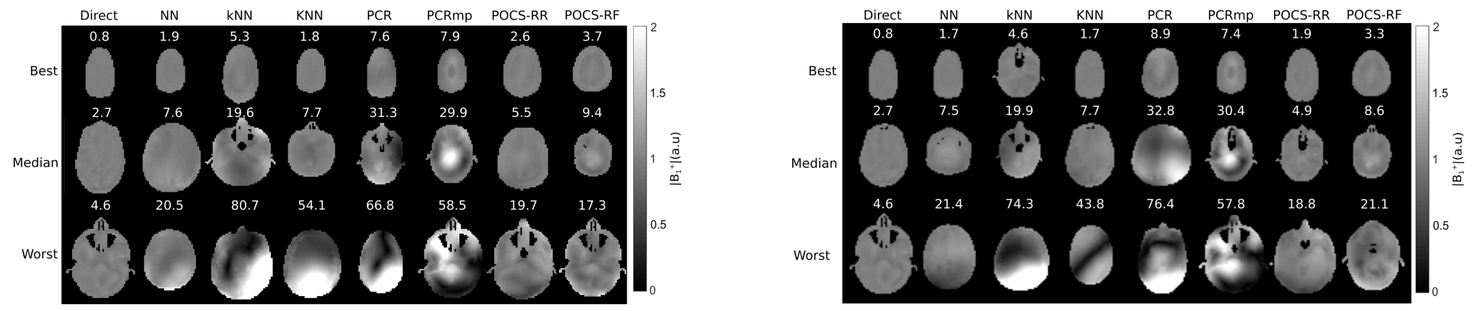


Figure 7. Best, median, and worst slice shims are shown for each learning method, compared to the results from the best directly designed (tailored) shims. The standard deviation of each excitation profile (as a percentage) is shown above each sample. A) shows the results with a B_1^+ map neighborhood of 3 and the sign features used, and b) shows results for a neighborhood of 1 with cross-term features.

Principal components regression (PCR/PCRmp), k-nearest neighbors (kNN) and kernelized NN (KNN) methods resulted in very non-uniform profiles. The POCS-RR, POCS-RF, and NN methods were the only .methods to improve uniformity over the circularly polarized (CP) mode (the best out-of-the-box solution). From Fig. 8 it is clear that direct design still performs significantly better than most of the learning methods; however, several methods perform decently, with POCS-RR achieving the best homogeneity overall.

The improvement of the POCS-RR method in Case 2 (Fig 8b) over Case 1 (Fig 8a) shows the importance of meaningful feature selection; a better result was achieved using *fewer raw B_1^+ map features* by introducing cross-terms and terminating the POCS shim training early to avoid overfitting. This resulted in more homogeneous profiles overall. However, the POCS-RF method performed slightly better in Case 1 than Case 2, emphasizing that there is no "one-size-fits all" optimal feature set across learning methods.

Excitation Uniformity

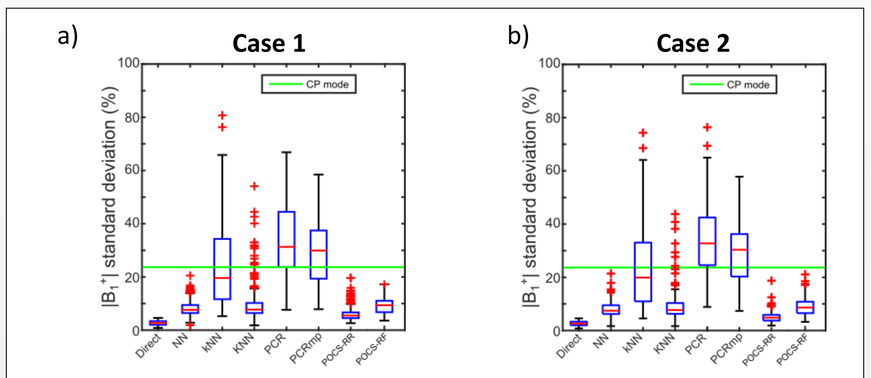


Figure 8. Standard deviation of the excitation profiles generated by the predicted shims, for each learning method; the same is shown for the profiles generated by direct design of shims. The green line indicates the median of the circularly polarized mode.

B_1^+ map Feature Analysis

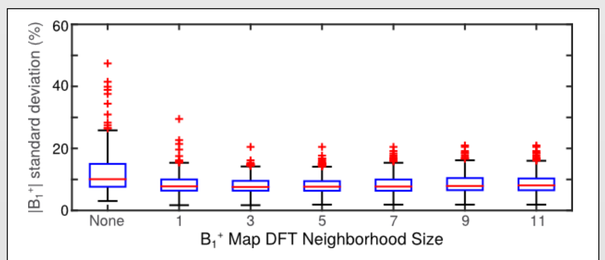


Figure 10. Standard deviation of the excitation profiles generated by the nearest-neighbor-predicted shims versus the neighborhood size of the Fourier transformed B_1^+ maps used for training.

The analysis of the neighborhood size of B_1^+ map Fourier coefficients necessary (for NN) shows that inclusion of a single coefficient from each coil helps shim performance of the nearest neighbors method greatly; after this, utility of including larger neighborhood sizes falls off.

Shim Weight Errors

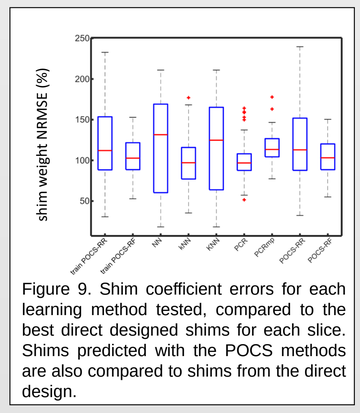


Figure 9. Shim coefficient errors for each learning method tested, compared to the best direct designed shims for each slice. Shims predicted with the POCS methods are also compared to shims from the direct design.

The normalized root-mean square errors (NRMSE) between the best directly designed and predicted shims (Fig 9) are very high, and not indicative of the degree of profile uniformity. These high errors are due to the non-convexity of the problem—this implies many solutions exist which produce similarly homogeneous fields, but have very different shim coefficients. This motivates the POCS methods, which enforce similarity in the target shims before interpolating over them.

DISCUSSION

The POCS shim methods performed best to predict RF shims. The gains in performance of predicted shims are substantial compared to other methods. Although here we have enforced predictability via ridge regression or random forests (2nd projection), this step could be replaced by other learning methods.

The learning methods that interpolate across training shims (kNN, KNN, PCR, PCRmp) fail because they aim to minimize the shim weight error instead of profile uniformity, which is the true target.

Shim performance of the POCS-RR method was comparable to that of direct design with the exception of several outliers. While POCS-RF performed slightly worse as a whole, the distribution of shim errors is tighter (case 1), sacrificing prediction accuracy for consistency.

RF shim prediction could also impact patient safety in the form of RF heating. Methods that impose a unit-sum weighted average over the training set boast an advantage in meeting specific absorption rate (SAR) constraints; if all of the solutions in the training set meet SAR constraints, then due to the convexity of the constraints, the predicted shims will also satisfy them. Several methods (NN, kNN, KNN) meet this requirement already, and POCS-RR should be possible to formulate this way also. This could obviate the need for SAR calculation for (predicted) RF shims at the scanner, and potentially enable more precise SAR constraints, since the computational burden is all at training time, not at the scanner.

CONCLUSION

Supervised learning methods can potentially save on time lost in the scanner for patient-tailored RF shimming, both in terms of scan time and compute time. The performance of the POCS-RR and -RF shim methods suggest an advantage of integrating the prediction method into shim design, rather than designing and predicting in series.

REFERENCES

[1] TS Ibrahim et al, MRI 19:1339-1347 (2001). [2] WMao et al, MRM 56:918-922 (2006). [3] R Lattanzi et al, MRM 61:315-334 (2009). [4] A Hoyos-Idrobo et al, IEEE TMI 33:739-748 (2014). [5] Christ et al., *Phys. Med. Biol.* 2010. [6] K Setsompop et al, MRM 59:908–915 (2008). [7] L. Breiman, Machine Learning 45:5–32 (2001).

ACKNOWLEDGEMENTS

The researchers acknowledge support NIH R01 EB016695. We thank Bennett Landman for helpful discussions.

Vanderbilt University Institute of Imaging Science
VUIIS Building, Medical Center North
1161 21st Ave, South
Nashville, TN 37232-2310
Tel: (615) 322-8359
Fax: (615) 322-0734
URL: www.vuiis.vanderbilt.edu