

Technical note: MAMBA—Multi-pAradigM voxel-Based Analysis: A computational cookbot

Giuseppe Palma¹ | Laura Cella² | Serena Monti²

¹Institute of Nanotechnology, National Research Council, Lecce, Italy

²Institute of Biostructures and Bioimaging, National Research Council, Napoli, Italy

Correspondence

Giuseppe Palma, Institute of Nanotechnology, National Research Council, Campus Ecotekne, Via Monteroni, 73100 - Lecce, Italy.
Email: giuseppe.palma@cnr.it

Abstract

Background: Voxel-Based (VB) analysis embraces a multifaceted ensemble of sophisticated techniques, lying at the boundary between image processing and statistical modeling, that allow for a frequentist inference of pathophysiological properties anchored to an anatomical reference. VB methods has been widely adopted in neuroimaging studies and, more recently, they are gaining momentum in radiation oncology research. However, the price for the power of VB analysis is the complexity of the underlying mathematics and algorithms.

Purpose: In this paper, we present the Multi-pAradigM voxel-Based Analysis (MAMBA) toolbox, which is intended for a flexible application of VB analysis in a wide variety of scenarios in medical imaging and radiation oncology.

Methods: The MAMBA toolbox is implemented in Matlab. It provides open-source functions to compute VB statistical models of the input data, according to a great variety of regression schemes, and to derive VB maps of the observed significance level, performing a non-parametric permutation inference. The toolbox allows for including VB and global outcomes, as well as an arbitrary amount of VB and global Explanatory Variables (EVs). In addition, the Matlab Parallel Computing Toolbox is exploited to take advantage of the perfect parallelizability of most workloads.

Results: The use of MAMBA was demonstrated by means of several realistic examples on a synthetic dataset mimicking a radiation oncology scenario.

Conclusion: MAMBA is an open-source toolbox, freely available for academic and non-commercial purposes. It is designed to make state-of-the-art VB analysis accessible to research scientists without the programming resources needed to build from scratch their own software solutions. At the same time, the source code is handed out for more experienced users to complement their own tools, also customizing user-defined models. MAMBA guarantees high generality and flexibility in the design of the statistical models, significantly expanding on the features of available free tools for VB analysis. The presented toolbox aims at increasing the reach of VB studies as well as the sharing of research results.

KEYWORDS

open-source software, radiation oncology, voxel-based analysis

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

1 | INTRODUCTION

Over the last three decades, different methods have been devised for investigating spatial relationships between several types of medical scalar fields and clinical conditions of patients. Such methods, overall referred to as Voxel-Based (VB) analysis, enable valuable inferences of pathophysiological properties spatially normalized to a Common Coordinate System (CCS). In particular, VB analysis has been widely used for assessing structural or functional changes in neuroanatomical images involving a voxel-wise comparison of the intensity (in Magnetic Resonance Imaging [MRI]) or activity (in nuclear medicine) of brain tissue between groups of individuals.^{1,2} VB morphometry represents a very useful tool for investigating the pathophysiological patterns of several neuroinflammatory or neurodegenerative diseases.³

Following the example of neuroimaging studies, VB analyses have been recently introduced in the field of radiation oncology. In this context, VB methods were applied to the analysis of dose distributions in which local dose response patterns are evaluated against a given radiation-related outcome. So far VB analyses have been indeed applied to several tumor types and to different toxicity endpoints^{4–7} or survival.^{8,9}

A VB analysis involves two main steps. First, patient images or maps are spatially normalized to a CCS defined by a template. Once images or maps required for a VB model have been accurately anchored, a statistical analysis is performed according to a given regression scheme to make inferences on group differences.

In this paper, we present a Matlab (MathWorks—Natick, Massachusetts, USA) toolbox for the application of multi-paradigm VB statistical analysis, called MAMBA, specifically designed to conform with some principles that reflect the authors' view on good practice in VB analysis: (1) to relax the parametric assumption on involved distributions; (2) to correct for massive multiple comparisons; and (3) to allow a Threshold-Free Cluster-Enhancement (TFCE) filtering of the statistic maps.¹⁰ In addition, MAMBA guarantees high generality and flexibility in the design of the statistical models, by bringing together for the first time: (1) built-in design of Generalized Linear Models (GLMs) and Cox proportional hazards regressions; (2) the opportunity to customize user-defined models; (3) arbitrarily mixed global and VB regressors/outcomes; and (4) a native parallel computing of the VB models.

This first release of the toolbox is open source, distributed under GPL license, and freely available for download at <https://github.com/pippipalma/MAMBA> for noncommercial research and educational uses.

The toolbox is implemented in Matlab in a procedural programming style to provide a non-proprietary solution to the scientific community users, who can freely adapt the functions according to their own needs. Mat-

lab, indeed, provides a simple syntax, as well as flexible high-level language constructs, good floating-point computational speed, and cross-platform availability. These features triggered a widespread diffusion of Matlab in the biomedical community, with a further positive feedback sparked by the popularity of some toolboxes externally developed by specific research groups and made freely available to the users (e.g., SPM12 in neuroimaging or CERR in radiation oncology¹¹).

In addition, the Parallel Computing Toolbox allows for a straightforward coding of high-level constructs that enable the parallelization of statement iterations in a loop on a parallel pool of workers on multi-core computers or clusters, without the need of programming with the Message Passing Interface. This represents an invaluable opportunity in the context of VB analysis, since both VB statistical models and significance permutation tests are examples of perfectly parallelizable workloads.

MAMBA runs on Matlab R2019b and later, and most common running configurations make extensive usage of the Statistics and Machine Learning Toolbox. Earlier Matlab releases might require minor changes on the code; while they could be easily implemented by experienced Matlab users, we might plan to extend the backward compatibility of the toolbox based on the community feedback.

2 | METHODS

The aim of MAMBA is to make the statistical segment of VB analysis accessible to scientific community, offering a unique set of statistical and computational features. It provides a comprehensive tool that, starting from clinical data and spatially normalized images, supports the user, with highly configurable built-in functions, through the data preparation for the VB statistical analysis, to the appropriate definition of advanced VB model and the computation of the associated statistics.

2.1 | Main functions

The MAMBA toolbox consists of four main functions:

1. `image_read`;
2. `image_clin_merge`;
3. `VBmodel`;
4. `perm_test`.

They are meant to be directly called by the user, as opposed to a set of auxiliary functions (stored in the subfolders `Engine/` and `External/`), which are required by the main functions to perform internal steps and should not be considered of direct interest to the end-user performing a VB analysis. Here, we introduce the tasks

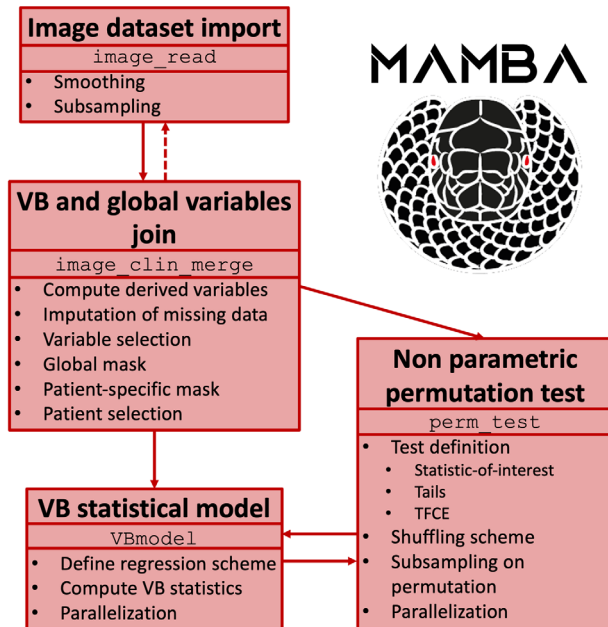


FIGURE 1 Flowchart of Multi-pAradigM Voxel-Based Analysis (MAMBA) toolbox. TFCE, Threshold-Free Cluster-Enhancement; VB, Voxel-Based.

performed by the four main functions and how they are meant to interact and be used (Figure 1).

The detailed description of the toolbox functions can be found in the Chapter “Reference manual” of Docs/UserManual.pdf on the GitHub repository.

The MAMBA toolbox revolves around one main function, `VBmodel`, which produces a VB statistical model of the input data according to several regression schemes. The built-in schemes include GLMs, Cox proportional hazards regressions, and analyses of variance (ANOVA); in addition, MAMBA allows for the design of customizable user-defined model. `VBmodel` takes as input the data associated with the cohort of patients to be modeled and returns as output a list of model statistics. The function is able to handle input data that could include both global (e.g., age, sex, etc.) and VB (e.g., spatially normalized ND images, with $N \geq 1$) variables, independently as regressors or outcome variables. `VBmodel` allows for a flexible storage of potentially many model statistics (e.g., GLM coefficients β , t statistics or p -values for β , etc.) for each voxel within a single VB array. The `VBmodel` function includes a large number of options that allow to specify, for example, the distribution of the outcome, as well as the parallelization strategy.

In a typical VB analysis, `VBmodel` is directly invoked by the user to inspect the effect size associated with each regressor of interest, for example, in terms of coefficients β in a GLM or hazard ratios in a Cox regression.

Also, `VBmodel` is invoked multiple times through the call to the second main function `perm_test`, which

derives a VB map of the observed significance level for a specific regressor of interest. It performs a non-parametric permutation inference¹² for the general class of statistical models suitable for `VBmodel`, including but not limited to the ones provided with MAMBA. To correct the significance map for the multiple comparisons, `perm_test` controls for the familywise error rate by estimating the null distribution of a summary statistic, which is given by the maximum VB statistic (hereafter `maxT`, regardless of the actual symbol for the statistic) over the region of interest.¹³ The input type signature of `perm_test` is the same of `VBmodel`, except for some additional options that define the hypothesis test associated with the VB model. For example, the function `perm_test` allows to configure the shuffling strategy of the permutation test (i.e., the “Shuffle Y” scheme^{12,14,15} or the “Shuffle Z” scheme^{12,15,16}), as well as the side of the test to perform. The TFCE filtering of the statistic maps can also be enabled and configured. In addition, the execution of `perm_test` can be speeded up using a subsampling scheme with negligible cost in terms of approximation error, as well as exploiting the implemented parallelization strategy.

There are several accepted formats for input data to `VBmodel` and `perm_test`, in order to ease the use of the MAMBA toolbox in both real and in silico analyses, as described in the toolbox reference manual. They include multi-dimensional numerical arrays for both Explanatory Variables (EVs) and outcome, as well as structures of data containing also patient-specific spatial masks. Such formats typically prove useful in studies performed in silico, in which the global variables characterizing the virtual patients within the VB model are lightly stored as numeric or logical values.

However, the global data on real cohorts of patients are typically stored in table-like structures, such as spreadsheet documents, which include demographic and clinical variables in potentially many data types (numbers, logicals, categorical variables, dates, time series, etc.). On the other side, the ND spatially normalized images or maps (morphological images, such as CT or MRI; functional images, such as PET or SPECT; regions-of-interest and structure masks; radiation dose spatial distributions or surface maps; etc.) required for the VB model are likely saved in patient-specific files produced by a co-registration tool. Therefore, in our experience, two additional functions help to handle the phase preceding a clinical VB analysis: `image_read` and `image_clin_merge`.

The function `image_read` reads the patient-specific maps according to a set of options customizable by the user, possibly including a smoothing step with a Gaussian kernel and a grid subsampling. It can be called directly by the user, or can be invoked by the function `image_clin_merge`, which handles the remaining data pre-processing before it is used by `VBmodel` and `perm_test`: it joins patient-wise the images obtained

by `image_read` and the variables in the table with demographic/clinical data; it computes the derived variables specified by the users according to a flexible and easy-to-implement Polish (i.e., prefix) notation; it provides a customizable framework for the imputation of missing data (for a careful handling of imputation technique please refer to Van Buuren et al.¹⁷); and, finally, it offers a selection tool of global EVs based on the elastic net regularization of the outcome regression.¹⁸ The outputs of the `image_clin_merge` function are a structure with general info on the dataset and a ready-to-use table. They contain all the required data for the following statistical analysis, in a format that allows users to easily set the EVs and the outcome of their VB model, without concern regarding their dimensionality. In particular, the `image_clin_merge` function is able to provide both a global region of interest, or a refined version of it according to user-defined operations, and patient-specific masks for the subsequent VB analysis. In addition, the function can perform a patient selection based on the conditions defined by the user on the available variables.

2.2 | User interaction

MAMBA is an intuitive command-line toolbox, conceived to allow also users without programming experience to design complex VB model and hypothesis test in a user-friendly way.

Indeed, the options of the four main MAMBA functions can be easily configured specifying pairs of name-value arguments; a scalar structure or a path to MAT-files, from which field name-value pairs are extracted. For each pair, a default value is set according to most common modes of use perceived by the authors of the toolbox.

The interested reader may refer to the MAMBA Reference Manual for a comprehensive description of the available options and on how to provide them.

A typical and complete VB statistical analysis on clinical data consists of three main steps in the MAMBA toolbox: the data reading by the function `image_clin_merge` (invoking `image_read`); the calculation of the voxel-wise effect size (e.g., the regression coefficient or the mean difference) of interest by the function `VBmodel`; and the computation of the associated p -map by the function `perm_test`.

2.3 | Input data

When starting a new analysis, MAMBA needs clinical/demographic data and images that are spatially normalized to a template. The `image_read` and the `image_clin_merge` functions are designed to read, respectively, images and clinical table from MAT-files. In particular, for spatially normalized images, this allows

to be independent from the specific format of the chosen pre-processing software. Users are invited to use existing third-party tools for importing the specific file types associated to the normalized images and the Matlab proprietary functions to import clinical spreadsheets.

For each patient, MAMBA requires a separate MAT-file containing the images that might be read and included into the analysis, together with patient-specific masks, if available. Also, time series of VB variables (e.g., diagnostic images acquired at several time-points or RT dose maps delivered in sequential treatments or re-irradiations), in which the time variable could play a relevant role, can be handled by MAMBA and stored in the patient MAT-file.

Finally, the anatomical structures that may be used to define the global region of interest for the subsequent VB analysis are expected to be stored in the template MAT-file.

3 | WORKED EXAMPLES AND RESULTS

On the GitHub repository, the user will find the Chapter “Worked examples and clinical results” in `Docs/UserManual.pdf`. Therein, we present some explicative results obtained using MAMBA on a purposely generated synthetic dataset, providing a direct means to test the proposed computational tool; we provide a full description of the toolbox options used in the worked-out examples, along with the toolbox configurations that led to clinical results previously published in the literature.^{19–23}

While MAMBA is intended for general-purpose VB analysis, the examples presented are drawn from scenarios related to radiation oncology, since the application of VB methods in this field is relatively recent compared to the experience of medical imaging and, specifically, neuroimaging. Therefore, researchers interested in exploring radiobiology patterns through VB analysis are likely to take particular advantage of dedicated examples.

4 | DISCUSSION AND CONCLUSION

In this paper, we presented the MAMBA toolbox, a set of Matlab functions for the application of multi-paradigm VB statistical analysis in the field of medical imaging and radiation oncology.

MAMBA represents the computational companion to our previously published VB analysis methodological cookbook,¹³ where the key elements and possible options for accurate VB statistical inference on dose maps were discussed, and the underlying mathematics was theoretically elucidated. In the computational cookbot introduced by the present paper, instead, we

TABLE 1 Comparison of features included in MAMBA and in other free tools for Voxel-Based (VB) analysis.

Feature in MAMBA	VoxelStats	PALM	SurfStat	BPM	glim_image
General linear model	✓	✓	✓	✓	✓
Generalized linear model	✓	X	X	X	✓
Cox regression	X	X	X	X	X
User-defined model	X	X	X	X	X
VB explanatory variables	✓	✓	X	✓	X
Global explanatory variables	✓	✓	✓	✓	✓
VB outcome	✓	✓	✓	✓	✓
Global outcome	✓	X	X	X	X
Interaction of VB explanatory variables	✓	X	X	X	X
Patient-specific masks	X	✓	X	X	X
Multiple comparison correction	✓	✓	✓	✓	X
Threshold-free cluster enhancement	X	✓	X	X	X
Non parametric statistics	X	✓	X	X	X
Parallel computing	✓	X	X	X	X

Abbreviation: BPM, Biological Parameter Mapping; MAMBA, Multi-pAradigm voxel-Based Analysis; PALM, Permutation Analysis of Linear Models.

implemented several strategies that we have devised while facing the problems encountered in many different clinical studies performed in the last years. In this way, MAMBA is intended to make the statistical segment of VB analysis accessible to radiation oncology community and to enable investigation on complex radiobiological mechanisms. At the same time, thanks to its high flexibility, it makes the methods originally developed for specific needs of the RT field available for the medical imaging community.

The set of statistical and computational features offered by MAMBA is, to the best of our knowledge, unique when compared with the similar open-source toolboxes,^{12,24–27} and substantially expands the scope of VB analysis applications (Table 1).

The great variety of available statistical models, along with the opportunity to include an arbitrary amount of VB or global EVs that impact on either VB or global outcomes, can open new ways of exploring statistical relationships in radiobiology²⁸ and multimodal imaging.²⁹ Such multi-variable VB analysis is likely to prove particularly valuable in quantitative MRI, which typically leads to the mapping of several, independent magnetic resonance parameters.^{30,31}

Clearly, the use of MAMBA requires a preprocessing phase consisting in an accurate spatial normalization step, which is not provided by our toolbox. Indeed, MAMBA is intended for general-purpose VB statistical analysis, while several and already optimized solutions exist for spatial normalization in different clinical settings, which are also standard *de facto* in the different communities of neuroimaging (e.g., SPM [<https://www.fil.ion.ucl.ac.uk/spm/>], FSL [<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>] and ANTs [<http://stnava.github.io/ANTs/>]) and radiation oncology (e.g., NiftyReg [<http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg>], Elastix [<http://elastix.isi.uu.nl>] and Plastimatch [<https://www.plastimatch.org/>]).

To facilitate its distribution, MAMBA is hosted at <https://github.com/pippipalma/MAMBA> as an open-source code freely available for academic and non-commercial purposes. Though we do not plan to establish a collaborative software development project, users are welcome to send comments and pull requests. In this respect, depending on the feedback from the users, we plan to implement in the future a graphical user interface; also, we expect to add new tools for the RT community related to the preliminary analysis of dose datasets^{21,32} and to the modeling of VB Normal Tissue Complication Probability (NTCP).³³

In conclusion, we hope that the MAMBA Toolbox will foster the diffusion of VB analyses as well as the consolidation of networks of researchers that could “fail better” and better in designing and applying VB tools.

ACKNOWLEDGMENTS

The authors are grateful to Massimiliano Sorbillo for his valuable assistance as logo designer. Amedeo Buonanno, Gian Paolo Marra, and Giovanni Trematerra are gratefully acknowledged for insightful discussions on software design patterns. The authors thank Sirio Coccozza, Giuseppe Pontillo, and Alfredo Montella for providing access to the NeuroN Lab server.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts to disclose.

REFERENCES

- Ashburner J, Friston KJ. Voxel-based morphometry—the methods. *Neuroimage*. 2000;11(6 Pt 1):805–821.

2. Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier; 2011.
3. Whitwell JL. Voxel-based morphometry: an automated technique for assessing structural changes in the brain. *J Neurosci*. 2009;29(31):9661-9664.
4. Acosta O, Dreon G, Ospina JD, et al. Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy. *Phys Med Biol*. 2013;58(8):2581-2595.
5. Palma G, Monti S, D'Avino V, et al. A voxel-based approach to explore local dose differences associated with radiation-induced lung damage. *Int J Radiat Oncol Biol Phys*. 2016;96(1):127-133.
6. Monti S, Pacelli R, Cella L, Palma G. Inter-patient image registration algorithms to disentangle regional dose bioeffects. *Sci Rep*. 2018;8(1):4915.
7. Beasley W, Thor M, McWilliam A, et al. Image-based data mining to probe dosimetric correlates of radiation-induced trismus. *Int J Radiat Oncol Biol Phys*. 2018;102(4):1330-1338.
8. McWilliam A, Kennedy J, Hodgson C, Vasquez Osorio E, Faivre-Finn C, van Herk M. Radiation dose to heart base linked with poorer survival in lung cancer patients. *Eur J Cancer*. 2017;85:106-113.
9. Green A, Vasquez Osorio E, Aznar MC, McWilliam A, van Herk M. Image based data mining using per-voxel cox regression. *Front Oncol*. 2020;10:1178.
10. Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*. 2009;44(1):83-98.
11. Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. *Med Phys*. 2003;30(5):979-985.
12. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. *Neuroimage*. 2014;92:381-397.
13. Palma G, Monti S, Cella L. Voxel-based analysis in radiation oncology: a methodological cookbook. *Phys Med*. 2020;69:192-204.
14. Manly BFJ. Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. *Res Popul Ecol (Kyoto)*. 1986;28(2):201-218.
15. Kennedy PE. Randomization tests in econometrics. *J Bus Econom Statist*. 1995;13(1):85-94.
16. Draper NR, Stoneman DM. Testing for the Inclusion of variables in linear regression by a randomisation technique. *Technometrics*. 1966;8(4):695-699.
17. Van Buuren S. *Flexible Imputation of Missing Data*. CRC press; 2018.
18. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67(2):301-320.
19. Monti S, Palma G, D'Avino V, et al. Voxel-based analysis unveils regional dose differences associated with radiation-induced morbidity in head and neck cancer patients. *Sci Rep*. 2017;7(1):7220.
20. Palma G, Monti S, Xu T, et al. Spatial dose patterns associated with radiation pneumonitis in a randomized trial comparing intensity-modulated photon therapy with passive scattering proton therapy for locally advanced non-small cell lung cancer. *Int J Radiat Oncol Biol Phys*. 2019;104(5):1124-1132.
21. Cella L, Monti S, Xu T, et al. Probing thoracic dose patterns associated to pericardial effusion and mortality in patients treated with photons and protons for locally advanced non-small-cell lung cancer. *Radiother Oncol*. 2021;160:148-158.
22. Monti S, Xu T, Liao Z, Mohan R, Cella L, Palma G. On the interplay between dosimics and genomics in radiation-induced lymphopenia of lung cancer patients. *Radiother Oncol*. 2022;167:219-225.
23. Monti S, Xu T, Mohan R, Liao Z, Palma G, Cella L. Radiation-induced esophagitis in non-small-cell lung cancer patients: voxel-based analysis and NTCP modeling. *Cancers*. 2022;14(7):1833.
24. Mathotaarachchi S, Wang S, Shin M, et al. VoxelStats: a MATLAB package for multi-modal voxel-wise brain image analysis. *Front Neuroinform*. 2016;10:20.
25. Worsley K, Liao C, Grabove M, Petre V, Ha B, Evans A. A general statistical analysis for fMRI data. *Neuroimage*. 2000;5(11):S648.
26. Casanova R, Srikanth R, Baer A, et al. Biological parametric mapping: a statistical toolbox for multimodality brain image analysis. *Neuroimage*. 2007;34(1):137-143.
27. https://github.com/BIC-MNI/glim_image
28. Palma G, Monti S, Conson M, Pacelli R, Cella L. Normal tissue complication probability (NTCP) models for modern radiation therapy. *Semin Oncol*. 2019;46(3):210-218.
29. Palma G, Tedeschi E, Borrelli P, et al. A novel multiparametric approach to 3D quantitative MRI of the brain. *PLoS One*. 2015;10(8):e0134963.
30. Monti S, Borrelli P, Tedeschi E, Coccozza S, Palma G. RESUME: turning an SWI acquisition into a fast qMRI protocol. *PLoS One*. 2017;12(12):e0189933.
31. Monti S, Pontillo G, Russo C, Cella L, Coccozza S, Palma G. RESUME(N): a flexible class of multi-parameter qMRI protocols. *Phys Med*. 2021;88:23-36.
32. Palma G, Monti S, Pacelli R, et al. Radiation pneumonitis in thoracic cancer patients: multi-center voxel-based analysis. *Cancers (Basel)*. 2021;13(14):3553.
33. Palma G, Monti S, Buonanno A, Pacelli R, Cella L. PACE: a probabilistic atlas for normal tissue complication estimation in radiation oncology. *Front Oncol*. 2019;9:130.

How to cite this article: Palma G, Cella L, Monti S. Technical note: MAMBA—Multi-pAradigM voxel-Based Analysis: A computational cookbot. *Med Phys*. 2023;50:2317–2322. <https://doi.org/10.1002/mp.16260>