An Open Source Multivariate Framework for n-Tissue Segmentation with Evaluation on Public Data

Brian B. Avants[1*], Nicholas J. Tustison[2*], Jue Wu[1], Philip A. Cook[1], and James C. Gee[1]

[1] Penn Image Computing and Science Laboratory, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

[2] Department of Radiology, University of Virginia, Charlottesville, Virginia, USA

Atropos: n-Tissue Segmentation
Corresponding author:
Brian B. Avants
3600 Market Street, Suite 370
Philadelphia, PA 19104
avants@picsl.upenn.edu

---

[*]The first two authors contributed equally to this work.

Abstract

We introduce *Atropos*, an ITK-based multivariate n-class open source segmentation algorithm distributed with ANTs.[1] The Bayesian formulation of the segmentation problem is solved using the Expectation Maximization (EM) algorithm with the modeling of the class intensities based on either parametric or non-parametric finite mixtures. Atropos is capable of incorporating spatial prior probability maps, (sparse) prior label maps and/or Markov Random Field (MRF) modeling. Atropos has also been efficiently implemented to handle large quantities of possible labelings (in the experimental section, we use up to 69 classes) with a minimal memory footprint. This work describes the technical and implementation aspects of Atropos and evaluates its performance on two different ground-truth datasets. First, we use the BrainWeb dataset from Montreal Neurological Institute to evaluate three-tissue segmentation performance via (1) K-means segmentation without use of template data; (2) MRF segmentation with initialization by prior probability maps derived from a group template; (3) Prior-based segmentation with use of spatial prior probability maps derived from a group template. We also evaluate Atropos performance by using spatial priors to drive a 69-class EM segmentation problem derived from the Hammers atlas from University College London. These evaluation studies, combined with illustrative examples that exercise Atropos options, demonstrate both performance and wide applicability of this new platform-independent open source segmentation tool.

---

[1]http://www.picsl.upenn.edu/ANTs

1 Introduction

As medical image acquisition technology has advanced, significant investment has been made towards adapting classification techniques for neuroanatomy. Early work appropriated NASA satellite image processing software for statistical classification of head tissues in 2-D MR images [1]. A proliferation of techniques ensued with increasing sophistication in both core methodology and degree of refinement for specific problems. The chronology of progress in segmentation may be tracked through both technical reviews [2–9] and evaluation studies [e.g. 10–13].

The problem of accurately delineating the white matter, grey matter and cerebrospinal fluid (and subdivisions) of the human brain continuously spurs technical development in segmentation. Following [1], many researchers adopted statistical methods for n-tissue anatomical brain segmentation. The Expectation-Maximization (EM) framework is natural [14] given the "missing data" aspect of this problem. The work described in [15] was one of the first to use EM for finding a locally optimal solution by iterating between bias field estimation and tissue segmentation. A core component of this work was explicit modeling of the tissue intensity values as normal distributions [16] for both 2-D univariate simulated data and T1 coronal images, which continues to find utility in contemporary developments. A secondary component was an extended non-parametric probability model, also influenced by earlier work [17], where Parzen windowing is used to model the tissue intensity distribution omitting consideration of the underlying bias field. Although technically not an EM-based algorithm, the robustness of the latter has motivated its continued use even more recently [e.g. 18].

Subsequent development included the use of Markov Random Field (MRF) modeling [19] to regularize the classification results [20] with later work adding heuristics concerning neuroanatomy to prevent over-regularization and the resulting loss of fine structural details [21, 22]. A more formalized integration of generic MRF spatial priors was employed in the work of [23], commonly referred to as FAST (FMRIB's Automated Segmentation Tool), which is in widespread use given its public availability and good performance. More recently, a uniform distribution of local MRFs within the brain volume and their subsequent integration into a global solution has been proposed obviating the need for an explicit bias correction solution [24].

Several initialization strategies have been proposed to overcome the characteristic susceptibility

of EM algorithms to local optima. Common low-level initialization steps include uniform probability assignment [15], Otsu thresholding [23], and K-means clustering [25]. More sophisticated low-level initialization schemes include that of [26] in which a dense spatial distribution of Gaussians is used to capture the complex neuroanatomical layout with subsequent processing used to conjoin subsets of such Gaussians belonging to the same tissue classes. Recently, reseachers have begun to rely on spatial prior probability maps of anatomical structures of interest to encode domain knowledge[22, 27, 28]. These spatial prior probability maps may also provide an initial segmentation. Related technological developments model partial volume effects for increased accuracy in brain segmentation [29–31].

A general trend towards more integrative neuroanatomical image processing led to the work described in [28] which is publicly available within SPM5, a large-scale Matlab module in which registration, segmentation, and bias field correction can be simultaneously modeled within a single optimization scheme. The roots of this very popular software package stem back to early work by Karl Friston which laid the basis for statistical parametric mapping [32]. Similar integrative brain processing was provided in [33] in which segmentation and registration parameters were optimized simultaneously while casting the inhomogeneity model parameters of [15] as nuisance variables. Continued work involved recursive parcellation of the brain volume by considering substructures in a hierarchical manner [34]. An implementation is provided in 3D slicer [35]—an open source medical image computation and visualization package with developmental contributions from multiple agencies including both private and academic institutions.

Related neuroanatomical research concerns the selection of geometric features of the cortex [e.g. 36] which aims at understanding the functional-anatomical relationship of the human brain. Recent endeavors produce a dense cortical labeling in which every point of the cortex is classified, i.e. a cortical parcellation [37–39]. Various techniques have been proposed to reduce the manual effort required to densely label a high-resolution neuroimage; one example is the popular software package known as Freesurfer [37, 40, 41]. In contrast to the volumetric approach detailed in this work, Freesurfer is primarily a surface-based technique in which the brain structures such as the grey-white matter interface and pial surfaces are processed, analyzed, and displayed as tessellated surfaces [40, 41]. Advantages of surface representations include the ability to map processed neuroanatomy to simple geometric primitives such as spheres or planes and the ease of including topological

constraints in the analysis workflow. These types of methods, including Klein's Mindboggle [42], would usually follow an initial segmentation by a volumetric method such as Atropos.

Researchers in aging often focus on accurately segmenting the T1 MRI of elderly controls and subjects suffering from neurodegeneration, for instance, via SIENA [43]. A recent evaluation study compared kNN segmentation, SPM Unified Segmentation and SIENA and found different performance characteristics under different evaluation criteria [44]. [12] had similar findings when comparing SPM5, FSL and FreeSurfer. These studies suggest that no single method performs best under every measurement and, along with the No Free Lunch theorem [45], highlight the need for segmentation tools that are tunable for different problems and research goals.

Our open source segmentation tool, which we have dubbed *Atropos*,[2] efficiently and flexibly implements an n-tissue paradigm for voxel-based image segmentation. Atropos allows users to harness its generalized EM algorithm for standard tissue classification of the brain into gray matter, white matter and cerebrospinal fluid even in cases of multivariate image data—relevant when more than one view of anatomy aids segmentation, as in neonatal brain tissue classification [e.g. 18, 46]. Atropos equally allows incarnations that use EM to simultaneously maximize the posterior probabilities of many classes with minimal random access memory requirements, for instance, when parcellating the brain into hemispheres, cortical regions and deep brain structures such as amygdala, hippocampus and thalamus. Atropos contains features of its predecessors for performing n-tissue segmentation including imposition of prior information in the form of MRFs and template-based spatial prior probability maps as well as weighted combinations of these terms. We also borrow an idea from [47] and use sparse spatial priors to provide initialization and boundary conditions for Atropos EM segmentation in a semi-interactive manner. In short, Atropos seeks to provide a segmentation toolbox that may be modified, tuned and refined for different use scenarios.

Coupled with the registration [48] and template building [49] already included in the ANTs, Atropos is a versatile and powerful software tool which touches multiple aspects of our brain processing pipeline. We use Atropos to address brain extraction [50], grey matter/white matter/cerebrospinal fluid segmentation, label fusion/propagation and cortical parcellation. We also allow Atropos to interact with the recently developed N4 bias correction software [51] in an adaptive manner. To

---

[2]Atropos is one of the three Fates from Greek mythology characterized by her dreaded shears used to decide the destiny of each mortal. Also, consistent with the entomological motif of our ANTs, *Acherontia atropos* is a species of large moth known for the skull-like pattern visible on its thorax.

further highlight the value of this open source contribution, we performed a search of software attributes on NITRC and found that as of November 2010 no stand-alone EM methods are currently listed. We also evaluate Atropos performance on two brain MRI segmentation objectives. First, three-tissue classification. Second, we test our ability to parcellate the brain into 69 neuroanatomical regions to illustrate the practical value of the low-memory implementation within this paper. Although Atropos may be applied to multivariate data from arbitrary modalities, we limit our evaluation to tissue classification in T1 neuroimaging in part due to the abundance of "gold-standard" data for this modality. Consistent with our advocacy of open science (not to mention the facilitation of analysis due to accessibility) we also only use publicly available data sets. For this reason, all results in this paper are reproducible with the caveat that users may require some guidance from the authors or other users in the community.

Organization of this work is as follows: we first describe the theory behind the various components of Atropos while acknowledging that more theoretical discussion is available elsewhere. This is followed by a thorough discussion of implementation which, though often overlooked, is of immense practical utility. We then report results on the BrainWeb and Hammers dataset. Finally, we provide a discussion of our results and our open source contribution in the context of the remainder of this paper and of previous and future work.

2 Theoretical Foundations for Atropos Segmentation

Atropos encodes a family of Bayesian segmentation techniques that may be configured in an application-specific manner. The theory underlying Atropos dates back 20+ years and is representative of some of the most innovative work in the field. Although we summarize some of the theoretical work in this section, we recommend that the interested reader consult the deep literature in this field for additional perspective and proofs behind the major concepts.

Bayes' theorem provides a powerful mechanism for making inductive inferences assuming the availability of quantities defining the relevant conditional probabilities, specifically the likelihood and prior probability terms. Bayesian paradigms for brain image segmentation employ a user selected observation model defining the likelihood term and one or more prior probability terms. The product of likelihood(s) and prior(s) is proportional to the posterior probability. The likelihood term has been previously defined both parametrically (e.g. a Gaussian model) and non-parametrically

(e.g. Parzen windowing of the sample histogram). The prior term, as given in the literature, has often been formed either as MRF-based and/or template-based. An image segmentation solution in this context is an assignment of one label to each voxel[3] such that the posterior probability is maximized. The next sections introduce notation and provide a formal description of three essential components in Bayesian segmentation, viz.

- the likelihood or observation model(s),

- the prior probability quantities derived from a generalized MRF and template-based prior terms, and

- the optimization framework for maximizing the posterior probability.

These components are common across most EM segmentation algorithms.

2.1   Notation

Assume a field, $\mathcal{F}$, whose values are known at discrete locations, i.e. sites, within a regular voxel lattice that makes up an image domain, $\mathcal{I}$. Note that $\mathcal{F}$ can be a scalar field in the case of unimodal data (e.g. T1 image only) or a vector field in the case of multimodal data (e.g. T1, T2, and proton density images). A specific set of observed values, denoted by $\mathbf{y}$, are indexed at N discrete locations in $\mathcal{I}$ by $i \in \{1, 2, \ldots, N\}$. This random field, $Y = \{y_1, y_2, \ldots, y_N\}$, serves as a discrete representation of an observed image's intensities. A labeling of this image, also known as a hard segmentation, assigns to each site in $\mathcal{I}$ one of K labels from the finite set $\mathcal{L} = \{l_1, l_2, \ldots, l_K\}$. Also considered a random field, this discrete labeling is $X = \{x_1, x_2, \ldots, x_N\}$ where each $x_i \in \mathcal{L}$. We use $\mathbf{x}$ to denote a specific set of labels in $\mathcal{I}$ and a valid, though not necessarily optimal, solution to the segmentation problem.

2.2   Segmentation Objective Function

Atropos optimizes a class of user selectable objective functions each of which may be represented in a generic Bayesian framework, as described by [52]. This framework requires *likelihood models*

---

[3]In the classic 3-tissue segmentation case, each voxel in the brain region is assigned a label of 'cerebrospinal fluid (csf)', 'gray matter (gm)', or 'white matter (wm)'.

and *prior models* which enter into Bayes' formula,

$$p(\mathbf{x}|\mathbf{y}) = \underbrace{p(\mathbf{y}|\mathbf{x})}_{\text{Likelihood(s)}} \underbrace{p(\mathbf{x})}_{\text{Prior(s)}} \frac{1}{p(\mathbf{y})} \tag{1}$$

where the normalization term, $1/\mathbf{y}$, is a constant that does not affect the optimization [52]. Given choices for likelihood models and prior probabilities, the Bayesian segmentation solution is the labeling $\hat{\mathbf{x}}$ which maximizes the posterior probability, i.e.

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}} \left\{ p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}) \right\}. \tag{2}$$

Similar to its predecessors, Atropos employs the EM framework [14] to find maximum likelihood solutions to this problem. The following sections detail the Atropos EM along with choices for the likelihood and prior terms.

2.3   Likelihood or Observation Models

To each of the K labels corresponds a single probabilistic model describing the variation of $\mathcal{F}$ over $\mathcal{I}$. We denote this set of K likelihood models as $\Phi = \{p_1, p_2, \ldots, p_K\}$. Using the standard notation, $\Pr(S = s) = p(s)$, $\Pr(S = s|T = t) = p(s|t)$, we can define these voxelwise probabilities, $\Pr_k(Y_i = y_i|X_i = l_k) = p_k(y_i|l_k)$, in either parametric or non-parametric terms. Given its simplicity and good performance, in the parametric case, p_k is typically defined as a normal distribution, i.e.

$$p_k\left(y_i|l_k\right) = G\left(\mu_k; \sigma_k\right)$$
$$= \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-(y_i - \mu_k)^2}{2\sigma_k^2}\right) \tag{3}$$

where the parameters μ_k and σ_k^2 respectively represent the mean and variance of the kth model. When y_i is a vector quantity, we replace the Euclidean distance by Mahalanobis distance and define multivariate Gaussian parameters via a mean vector, $\boldsymbol{\mu}_k$, and covariance matrix, $\boldsymbol{\Sigma}_k$.

A common technique for the non-parametric variant is to define p_k using Parzen windowing of

the sample observation histogram of $\mathbf{y}$, i.e.

$$p_k\left(y_i|l_k\right) = \frac{1}{N_B}\sum_{j=1}^{N_B}\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left(\frac{-(y_i - c_j)^2}{2\sigma_j^2}\right)$$

(4)

where N_B is the number of bins used to define the histogram of the sample observations (in Atropos the default is $N_B = 32$) and c_j is the center of the jth bin in the histogram. σ_j is the width of each of the N_B Gaussian kernels. For multi-modal data in which the number of components of y_i is greater than one, a Parzen window function is constructed for each component. The likelihood value is determined by the joint probability given by their product.

Atropos segmentation likelihood estimates are based on the classical finite mixture model (FMM). FMM assumes independency between voxels to calculate the probability associated with the entire set of observations, $\mathbf{y}$. Spatial interdependency between voxels is modeled by the prior probabilities discussed in the next section. Marginalizing over the set of possible labels, $\mathcal{L}$, leads to the following probabilistic formulation

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{N}\left(\sum_{k=1}^{K}\gamma_k p_k(y_i|l_k)\right)$$

(5)

where γ_k is the mixing parameter [28].

2.4   Prior Probability Models

By modeling $\mathcal{F}$ via the set of observation models Φ, this so called finite-mixture model could be used to produce a labeling or segmentation [e.g. 15]. However, as pointed out by [23], exclusive use of the intensity profile produces a less than optimal solution because spatial contextual considerations are ignored. This has been remedied by the introduction of a host of prior probability models including those characterized by use of MRF theory and template-based information. For example, in the works of [22] and [18], the original global prior term given in [15] is replaced by the product of the template-based and the MRF-based prior terms. In addition to their descriptions below, we discuss a third possible prior/objective combination in the form of a (sparse) prior labeling which fixes specific points of the segmentation and uses EM to propagate this information elsewhere in the image.

2.4.1 Generalized MRF Prior

One may incorporate spatial coherence into the segmentation by favoring labeling configurations in which voxel neighborhoods tend towards homogeneity. This intuition is formally described by MRF theory in which spatial interactions in voxel neighborhoods can be modeled [53].

We assume the random field introduced earlier, X, is an MRF characterized by a neighborhood system, $\mathcal{N}_i$, on the lattice, $\mathcal{I}$, composed of the neighboring sites of i. This neighborhood system is both noninclusive, i.e. $i \notin \mathcal{N}_i$, and reciprocating, i.e. $i \in \mathcal{N}_j \Leftrightarrow j \in \mathcal{N}_i$. As an MRF, X also satisfies the positivity and locality conditions, i.e.,

$$p(\mathbf{x}) > 0, \ \forall \mathbf{x} \tag{6}$$

and where $\mathbf{x}$ is any particular labeling configuration on X (in other words, any labeling permutation on X is *a priori* possible). The MRF locality condition is then,

$$p\left(x_i | x_{\mathcal{I}-\{i\}}\right) = p\left(x_i | x_{\mathcal{N}_i}\right) \tag{7}$$

where $x_{\mathcal{I}-\{i\}}$ is the labeling of the entire image lattice except at site i and $x_{\mathcal{N}_i}$ is the labeling of $\mathcal{N}_i$. This locality property enforces solely local considerations based on the neighborhood system in calculating the probability of the particular configuration, $\mathbf{x}$. Following these two assumptions, the Hammersley–Clifford theorem provides the basis for treating the MRF distribution (cf. Eqn. (6)) as a Gibbs distribution [19, 54], i.e.

$$p(\mathbf{x}) = Z^{-1} \exp\left(-U(\mathbf{x})\right) \tag{8}$$

with Z a normalization factor known as the *partition function* and $U(\mathbf{x})$ the *energy function* which can take several forms [53]. In Atropos, as is the case with many other segmentation algorithms of the same family, we choose $U(\mathbf{x})$ such that it is only composed of a sum over pairwise interactions

between neighboring sites across the image,[4] i.e.

$$U(\mathbf{x}) = \beta \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_i} V_{ij}(x_i, x_j) \tag{10}$$

where V_{ij} is typically defined in terms of the Kronecker delta, δ_{ij}, based on the classical Ising potential (also known as a Potts model) [54]

$$V_{ij}(x_i, x_j) = \delta_{ij}$$
$$= \begin{cases} 0 & \text{if } x_i = x_j \\ 1 & \text{otherwise} \end{cases} \tag{11}$$

and β is a granularity term which weights the contribution of the MRF prior on the segmentation solution. Since Atropos allows for non-uniform neighborhood systems and systems in which not just the immediate face-connected neighbors are considered, we use the modified function also used in [55], which weights the interaction term by the Euclidean distance, d_{ij}, between interacting sites i and j such that

$$V_{ij} = \frac{\delta_{ij}}{d_{ij}} \tag{12}$$

so that sites in the neighborhood closer to i are weighted more heavily than distant sites.

2.4.2 Template-Based Priors

A number of researchers have used templates to both ensure spatial coherence and incorporate prior knowledge in segmentation. A common technique is to select labeled subjects from a population from which a template is constructed [e.g. 49, which is also available in ANTs]. Each labeling can then be warped to the template where the synthesis of warped labeled regions produces a prior probability map or prior label map encoding the spatial distribution of labeled anatomy which

---

[4]Using a more expansive definition of $U(\mathbf{x})$,

$$U(\mathbf{x}) = \sum_{i=1}^{N} \left(V_i(x_i) + \beta \sum_{j \in \mathcal{N}_i} V_{ij}(x_i, x_j) \right) \tag{9}$$

would permit casting the other prior terms inside the definition of $U(\mathbf{x})$ in the form of the external field $V_i(x_i)$ but, for clarity purposes, we consider them separately.

can be harnessed in joint segmentation/registration or Atropos/ANTs hybrids involving unlabeled subjects.

We employ the strategy given in [28] in which the stationary mixing proportions, $\Pr(x_i = l_k) = \gamma_k$ (cf. Eqn. (5)), describing the prior probability that label l_k corresponds to a particular voxel, regardless of intensity, are replaced by the following spatially varying mixing proportions,

$$\Pr(x_i = l_k) = \frac{\gamma_k t_{ik}}{\sum_{j=1}^{K} \gamma_j t_{ij}}. \tag{13}$$

The t_{ik} is the prior probability value at site i which was mapped, typically by image registration, to the local image from a template data set. The user may also choose mixing proportions equal to

$$\Pr(x_i = l_k) = \frac{t_{ik}}{\sum_{j=1}^{K} t_{ij}} \tag{14}$$

via the command line interface to the posterior formulation.

2.4.3 Supervised Semi-Interactive Segmentation

Brain segmentation methods have relied on user interaction for many years [56–59]. Atropos is capable of benefitting from user knowledge via an initialization and optimization that depends upon a spatially varying prior label image passed as input. Rapid, sparse labeling—with visualization provided by ITK-SNAP (www.itksnap.org)—enables an interaction and execution processing loop that can be critical to solving segmentation problems with challenging clinical data in which automated approaches fail. This part of Atropos design is inspired by the interactive graph cuts pioneered by [60] and which has spawned many follow-up applications. The Atropos prior label image pre-specificies the segmentation results at a subset of the spatial domain by fixing the priors and likelihood (and, thus, the posterior) at a subset of $\mathcal{I}$ to be 1 for the known label and 0 for each other label at the same site. The user input therefore not only initializes the optimization, but also gives boundary conditions that influence the EM solution outside of the known sites. While the graph-based min-cut max-flow solution is globally optimal for two labels, only locally optimal optimizers are available for 3 or more classes. Thus, in most practical applications, EM is a reasonable and efficient alternative to Boykov's solution. Furthermore, one may automate the initialization process. We provide this capability to allow the user to implement an interactive editing and seg-

mentation loop. The user may run Atropos with sparse manual label guidance, evaluate the results, update the manual labels and repeat until achieving the desired outcome. This processing loop be performed easily with, e.g., ITK-SNAP.

2.5 Optimization

Atropos uses expectation maximization to find a locally optimal solution for the user selected version of the Bayesian segmentation problem (cf. Eqn. (1)). After initial estimation of the likelihood model parameters, EM iterates between calculation of the missing optimal labels $\hat{\mathbf{x}}$ and subsequent re-estimation of the model parameters by maximizing the expectation of the complete data log-likelihood (cf. Eqn. (5)). The expectation maximization procedure is derived in various publications including [23] which yields the optimal mean and variance (or covariance), but sets the mixing parameter γ_k as a constant. The Atropos implementation estimates γ_k at each iteration, similar to [28].[5] When spatial coherence constraints are included as an MRF prior in Atropos, the optimal segmentation solution becomes intractable.[6] Although many optimization techniques exist (see the introduction in [27] for a concise summary of the myriad optimization possibilities)—each with their characteristic advantages and disadvantages in terms of computational complexity and accuracy—Atropos uses the well-known *Iterated Conditional Modes* (ICM) [61] which is greedy, computationally efficient and provides good performance. The EM employed in Atropos may therefore be written as a series of steps:

Initialization: In all cases, the user defines the number of classes to segment. The simplest initialization is by the classic K-means or Otsu thresholding algorithms with only the number of classes specified by the user. Otherwise, the user must provide prior information for each class in the form of either a single *n*-ary prior label image or a series of prior probability images, one for each class. The initialization also provides starter parameters.

Label Update (E-Step): Given the initialization and fixed model parameters, Atropos is capable of updating the current label estimates using either a *synchronous* or *asynchronous* scheme. The former is characterized by iterating through the image and determining which label

---

[5]Due to the lack of parameters in the non-parametric approach, it is not technically an EM algorithm (as described in [15]). However, the same iterative maximization is applicable and is quite robust in practice as evidenced by the number of researchers employing non-parametric models (see the Introduction).

[6]Consider N sites each with a possible K labels for a total of N^K possible labeling configurations. For large $K \gg 3$, exact optimization is even more intractable than for the traditional 3-tissue scenario.

maximizes the posterior probability without updating any labels until all voxels in the mask have been visited at which point all the voxel labels are updated simultaneously (hence the descriptor "synchronous"). This option is specified with `--icm [0]`. However, unlike asynchronous schemes characteristic of ICM, synchronous updates lack convergence guarantees. To determine the labeling which maximizes the posterior probability for the asynchronous option, an "ICM code" image is created once for all iterations by iterating through the image and assigning an ICM code label to each voxel in the mask such that each MRF neighborhood has a non-repeating code label set. Thus each masked voxel in the ICM code image is assigned a value in the range $\{1, \ldots, C\}$ where C is the maximum code label. Such an image can be created and viewed with Atropos by assigning a valid filename in the `--icm [1]` set of options. An example adult brain slice and the associated code image is given in Figure 1 for an MRF neighborhood of 5×5 pixels. This produces a maximum code label of '13'. For each iteration, a random ordering of the set $\{1, \ldots, C\}$ is created which prescribes the order in which the voxel labels are updated asynchronously. After the first pass through the set of code labels, additional passes can further increase the posterior probability until convergence (in ∼5 iterations). One can specify a maximum number of these "ICM iterations" on the command line. For our example in Figure 1, this means that for each ICM iteration, we would iterate through the image 13 times only updating those segmentation labels associated with the current ICM code.

Parameter Update (M-Step): Note that the posteriors used in the previous iteration are used to estimate the parameters at the current iteration. We use a common and elementary estimate of the mixing parameters:

$$\gamma_k \leftarrow \frac{1}{N} \sum_{i=1}^{N} p_k(l_k|y_i). \tag{15}$$

We update the parametric model parameters by computing, for each of K labels, the mean,

$$\mu_k \leftarrow \frac{\sum_{i=1}^{N} y_i p_k(l_k|y_i)}{\sum_{i=1}^{N} p_k(l_k|y_i)} \tag{16}$$
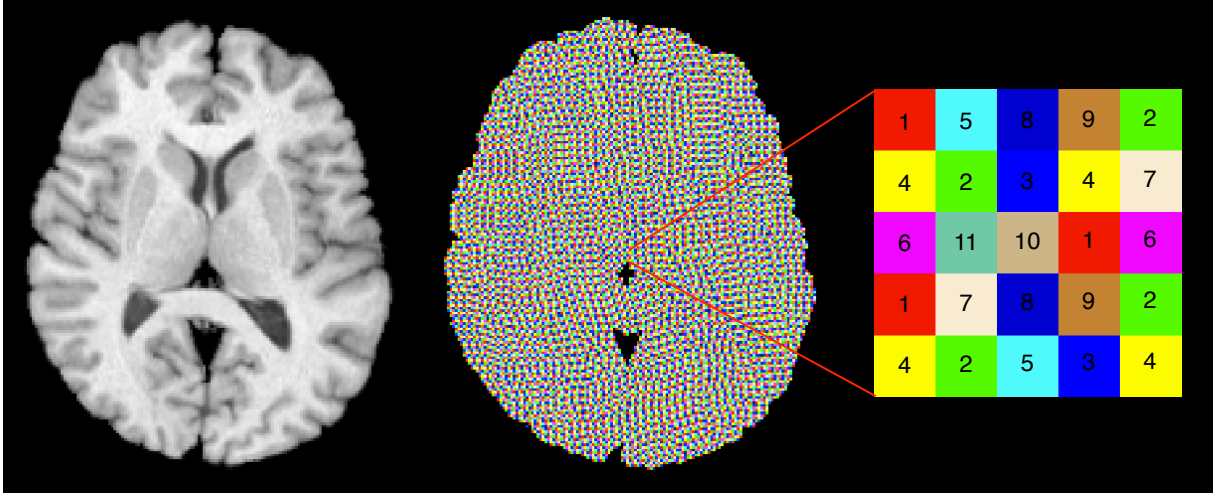
Figure 1: An adult brain image slice is shown with its ICM code image corresponding to a 5×5 MRF neighborhood. To the right of the ICM code image, we focus on a single neighborhood with a center voxel associated with the ICM code label of '10'. Each center voxel in a specified neighborhood exhibits a unique ICM code label which does not appear elsewhere in its neighborhood. When performing the segmentation labeling update for ICM, we iterate through the set of ICM code labels and, for each code label, we iterate through the image and update only those voxels associated with the current code label.

and variance,

$$\sigma_k^2 \leftarrow \frac{\sum_{i=1}^{N}(y_i - \mu_k)^T p_k(l_k|y_i)(y_i - \mu_k)}{\sum_{i=1}^{N} p_k(l_k|y_i)}. \tag{17}$$

The latter two quantities are modified, respectively, in the case of multivariate data as follows:

$$\boldsymbol{\mu}_k \leftarrow \frac{\sum_{i=1}^{N} \mathbf{y}_i p_k(l_k|\mathbf{y}_i)}{\sum_{i=1}^{N} p_k(l_k|\mathbf{y}_i)} \tag{18}$$

and the kth covariance matrix, $\boldsymbol{\Sigma}_k$, is calculated from

$$\boldsymbol{\Sigma}_k \leftarrow \frac{\sum_{i=1}^{N} p_k(l_k|\mathbf{y}_i)(\mathbf{y}_i - \boldsymbol{\mu}_k)^{\mathrm{T}}(\mathbf{y}_i - \boldsymbol{\mu}_k)}{1 - \sum_{i=1}^{N} p_k^2(l_k|\mathbf{y}_i)}. \tag{19}$$

This type of update is known as *soft* EM. Hard EM, in contrast, only uses sites containing label l_k to update the parameters for the kth model. A similar pattern is used in non-parametric cases.

EM will iterate toward a local maximum. We track convergence by summing up the maximum posterior probability at each site over the segmentation domain. The E-step, above, depends upon
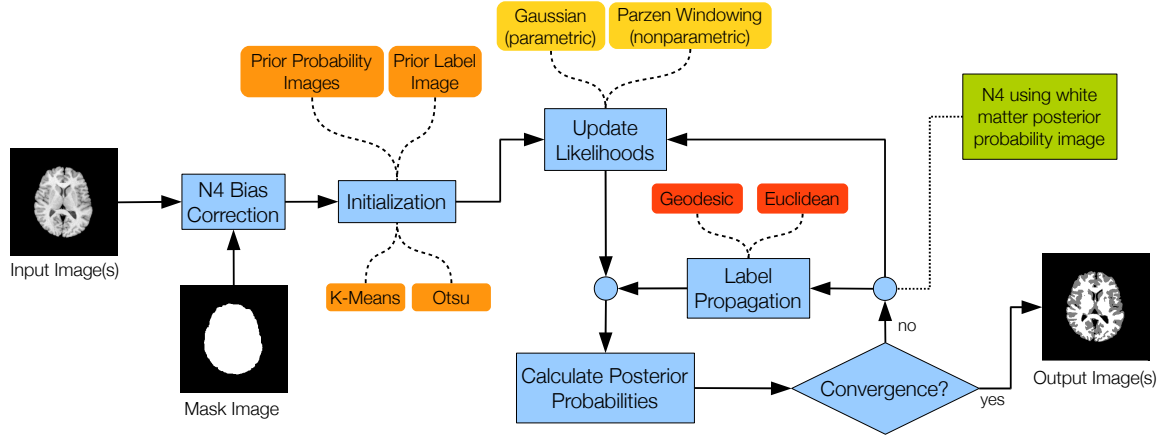
Figure 2: Flowchart illustrating Atropos usage typically beginning with bias correction via N4. Initialization provides an estimate before the iterative optimization in which the likelihood models for each class are tabulated from the current estimate followed by a recalculation of the posterior probabilities associated with each class. The multiple options associated with the different algorithmic components are indicated by the colored rounded rectangles connected to their respective core Atropos processes via curved dashed lines.

the selected coding strategy [61]. Atropos may use either a classical, sequential checkerboard update or a synchronous update of the labels, the latter of which is commonly used in practice. Synchronous update does not guarantee convergence but we employ it by default due to its intrinsic parallelism and speed. The user may alternatively select checkerboard update if he or she desires theoretical convergence guarantees. However, we have not identified performance differences, relative to ground truth, that convince us of the absolute superiority of one approach over the other.

3   Implementation

Organization of the implementation section roughly follows that of the theory section.

3.1   The Atropos User Interface

As with other classes that comprise ANTs, Atropos uses the Insight Toolkit as a developmental foundation. This allows us to take advantage of the mature portions of ITK (e.g. image IO) and ensures the integrity of the ancillary processes such as those facilitated by the underlying statistical framework. Although Atropos is publicly distributed with the rest of the ANTs package, we plan to contribute its core elements to the Insight Toolkit where it can be vetted and improved by other interested researchers.

An overview of Atropos components can be gleaned, in part, from the flowchart depicted in

Fig. 2. Given a set of input images and a mask images, each is preprocessed using N4 to correct for intensity inhomogeneity. For our brain processing pipeline, the mask is usually obtained from the standard skull-stripping preprocessing step which also uses Atropos. Initialization can be performed in various ways using standard clustering techniques, such as K-means, to prior-based images. This initialization is used to provide the initial estimate of the parameters of the likelihood model for each class. These likelihoods combine with the current labeling to generate the current estimate of the posterior probabilities at each voxel for each class. At each iteration, one can also integrate N4 by using the current posterior probability estimation of the white matter to update the estimate of bias field.

To provide a more intuitive interface without the overhead costs of a graphical user interface, a set of unique command line parsing classes were developed which can also provide insight to the functionality of Atropos. The short version of the command line help menu is given in Listing 1 which is invoked by typing 'Atropos -h' at the command prompt. Both short and long option flags are available and each option has its own set of possible values and parameters introduced in a more formal way in both the previous discussion and related papers cited in the introduction. Here we describe these options from the unique perspective of implementation.

3.2 Initializing the Atropos Objective

Atropos has a number of parameters defined within Listing 2 and will function on 2, 3 or 4 dimensional data. However, the majority of the time, users will be concerned with a smaller set of input parameters. Here, we list the recommended input and an example definition for each parameter:

Input images to be segmented: If more than one input image is passed, then a multivariate model will be instantiated. E.g. `-a Image.nii.gz` for one image and `-a Image1.nii.gz -a Image2.nii.gz` for multiple images.

Input image mask: This binary image defines the spatial segmentation domain. Voxels outside the masked region are designated with the label 0. E.g. `-x mask.nii.gz`.

Convergence criteria: The algorithm terminates if it reaches the maximum number of iterations or produces a change less than the minimum threshold change in the posterior. E.g. `-c [5,1.e-5]`.

```
COMMAND:
     Atropos

OPTIONS:
     -d, --image-dimensionality 2/3/4
     -a, --intensity-image [intensityImage,<adaptiveSmoothingWeight>]
     -b, --bspline [<numberOfLevels=6>,<initialMeshResolution=1x1x...>,<splineOrder=3>]
     -i, --initialization Random[numberOfClasses]
                          KMeans[numberOfClasses]
                          Otsu[numberOfClasses]
                          PriorProbabilityImages[numberOfClasses,
                            fileSeriesFormat(index=1 to numberOfClasses) or vectorImage,
                            priorWeighting,<priorProbabilityThreshold>]
                          PriorLabelImage[numberOfClasses,labelImage,priorWeighting]
     -p, --posterior-formulation Socrates[<useMixtureModelProportions=1>,
                                    <initialAnnealingTemperature=1>,<annealingRate=1>,
                                    <minimumTemperature=0.1>]
                                 Plato[<useMixtureModelProportions=1>,
                                    <initialAnnealingTemperature=1>,<annealingRate=1>,
                                    <minimumTemperature=0.1>]
     -x, --mask-image maskImageFilename
     -c, --convergence [<numberOfIterations=5>,<convergenceThreshold=0.001>]
     -k, --likelihood-model Gaussian
                            HistogramParzenWindows[<sigma=1.0>,<numberOfBins=32>]
     -m, --mrf [<smoothingFactor=0.3>,<radius=1x1x...>]
     -g, --icm [<useAsynchronousUpdate=1>,<maximumNumberOfICMIterations=1>,
                <icmCodeImage=''>]
     -o, --output [classifiedImage,<posteriorProbabilityImageFileNameFormat>]
     -u, --minimize-memory-usage (0)/1
     -w, --winsorize-outliers BoxPlot[<lowerPercentile=0.25>,<upperPercentile=0.75>,
                                  <whiskerLength=1.5>]
                              GrubbsRosner[<significanceLevel=0.05>,<winsorizingLevel=0.10>]
     -e, --use-euclidean-distance (0)/1
     -l, --label-propagation whichLabel[sigma=0.0,<boundaryProbability=1.0>]
     -h
     --help
```

Listing 1: Atropos short command line menu which is invoked using the '-h' option. The expanded menu, which provides details regarding the possible parameters and usage options, is elicited using the '--help' option.

MRF prior: The key parameter to increase or decrease the spatial smoothness of the label map is β. A useful range of β values is 0 to 0.5 where we usually use 0.05, 0.1 or 0.2 in brain segmentation. E.g. `-m [0.1, 1x1x1]` would define $\beta = 0.1$ with a MRF radius of one voxel in each of three dimensions.

Initialization: The initialization options include (where the first parameter defines K, here 3 for each below),

- `-i Kmeans[3]` standard K-means initialization for three classes,

- `-i PriorLabelImage[3,label_image.nii.gz]` and

- `-i PriorProbabilityImages[3,label_prob%02d.nii.gz,w]` where $w = 0$ (use the prior probability images only for initialization) or $w > 0.0$ (use the prior probability images throughout the optimization). If one chooses $0 < w < 1.0$ then one will increase (from zero) the weight on the priors. These images, like the `PriorLabelImage`, should be defined with the same domain as the input images to be segmented.

Posterior formulation: The user may choose to estimate the mixture proportions (or not) by setting `-p Socrates[1]` or `-p Socrates[0]`. Fixed label boundary conditions may be employed by selecting the `PriorLabelImage` initialization and `-p Plato[0]`.

Output: Atropos will output the hard segmentation and the probability image for each model. E.g. `-o [segmentation.nii.gz,seg_prob%02d.nii.gz]` will write out the hard segmentation in the first output parameter and a probability image for each class named, here, seg_prob01.nii.gz, seg_prob02.nii.gz, etc.

Higher dimensions than 4 are possible although we have not yet encountered such an application-specific need. Multiple images (assumed to be of the same dimension, size, origin, etc.), will automatically enable multivariate likelihoods. In that case, the first image specified on the command line is used to initialize the Random, Otsu, or K-means labeling with the latter initialization refined by incorporating the additional intensity images, i.e. an initial univariate K-means clustering is determined from the first intensity image which, along with the other images, provides the starting multivariate cluster centers for a follow-up multivariate K-means labeling. More details on each of the key implementation options are given below.

3.3   Likelihood Implementation

As mentioned previously in the introduction, different groups have opted for different likelihood models which have included either parametric (Gaussian) or non-parametric variations. However, these approaches are similar in that they require a list sample of intensity data from the input image(s) and a list of weighting values for each observation of the list sample from which the model is constructed. In general, one may query model probabilities by passing a given pixel's single intensity (for univariate segmentation) or multiple intensities (for multivariate segmentation) to the modeling function, regardless of whether the function is parametric or non-parametric. These similarities permitted the creation of a generic plug-in architecture where classes describing both parametric and non-parametric observational models are all derived from an abstract list sample function class. Three likelihood classes have been developed, one parametric and two non-parametric, and are available for usage although one of the non-parametric classes is still in experimental development. The plug-in architecture even permits mixing likelihood models with different classes during the same run for a hybrid parametric/non-parametric model although this possibility has yet to be fully explored.

If the Gaussian likelihood model is chosen, the list sample of intensity values and corresponding weights comprised of the posterior probabilities are used to estimate the Gaussian model parameters, i.e. the mean and variance. For the non-parametric model, the list sample and posteriors are used in a Parzen windowing scheme on a weighted histogram to estimate the observational model [62].

3.4   Prior Probability Models

3.4.1   Label Regularity

Consistent with our previous discussion, we offer both an MRF-based prior probability for modeling spatial coherence and the possibility of specifying a set of prior probability maps or a prior label map with the latter extendable to creating a dense labeling. To invoke the MRF '`-m/--mrf`' option, one specifies the smoothing factor (or the granularity parameter, β, given in Eqn. (10), and the radius (in voxels) of the neighborhood system using the vector notation '`1x1x1`' for a neighborhood radius of 1 in all 3 dimensions. This radius is defined such that voxels including but not limited to

those that are face-connected will influence the MRF.

3.4.2 Registration and Probability Maps

Image registration enables one to transfer information between spatial domains which may aid in both segmentation and bias correction. We rely heavily on template-building strategies [49, 50] which are also offered in ANTs. Since aligned prior probability images and prior label maps are often associated with such templates, Atropos can be initialized with these data with their influence regulated by a prior probability weighting term. Although prior label maps can be specified as a single multi-label image, prior probability data are often represented as multiple scalar images with a single image corresponding to a particular label. For relatively small classifications, such as the standard 3-tissue segmentation (i.e. white matter, gray matter, and cerebrospinal fluid), this does not typically present computational complexities using modern hardware. However, when considering dense cortical parcellations where the number of labels can range upwards of 74 per hemisphere [39], the memory load can be prohibitive if all label images are loaded into run-time memory simultaneously. A major part of minimizing memory usage in Atropos, which corresponds to the boolean '`-u/--minimize-memory-usage`' option, is the sparse representation of each of the prior probability images. Motivated by the observation that these spatial prior probability maps tend to be highly localized for large quantities of cortical labels, a threshold is specified on the command line (default = 0.0) and only those probability values which exceed that threshold are stored in the sparse representation. During the course of optimization, the prior probability image for a given label is reconstructed on the fly as needed. For instance, the NIREP (www.nirep.org) evaluation images are on the order of $300 \times 300 \times 256$ with 32 cortical labels. Our novel memory minimizing image representation typically shrinks run-time memory usage from a peak of 10+ GB to approximately 1.5 GB and enable these datasets to be used for training/prior-based cortical parcellation.

3.4.3 Integrating N4 Bias Correction

Assumptions about bias correction may be thought of as another prior model. As such, the typical segmentation processing pipeline begins with an intensity normalization/bias correction step using a method such as the recently developed N4 algorithm [51]. N4 extends the popular nonparametric

```
COMMAND:
     N4BiasFieldCorrection

OPTIONS:
     -d, --image-dimensionality 2/3/4
     -i, --input-image inputImageFilename
     -x, --mask-image maskImageFilename
     -w, --weight-image weightImageFilename
     -s, --shrink-factor 1/2/3/4/...
     -c, --convergence [<numberOfIterations=50>,<convergenceThreshold=0.001>]
     -b, --bspline-fitting [splineDistance,<splineOrder=3>,<sigmoidAlpha=0.0>,
                            <sigmoidBeta=0.5>]
                           [initialMeshResolution,<splineOrder=3>,<sigmoidAlpha=0.0>,
                            <sigmoidBeta=0.5>]
     -t, --histogram-sharpening [<FWHM=0.15>,<wienerNoise=0.01>,<numberOfHistogramBins=200>]
     -o, --output [correctedImage,<biasField>]
     -h
     --help
```

Listing 2: N4 short command line menu which is invoked using the '-h' option. The expanded menu, which provides details regarding the possible parameters and usage options, is elicited using the '--help' option.

nonuniform intensity normalization (N3) algorithm [63] in two principal ways:

- We replace the least squares B-spline fitting with a parallelizable alternative (which we also made publicly available in the Insight Toolkit)— the advantages being that 1) computation is much faster and 2) smoothing is not susceptible to outliers as is characteristic with standard least squares fitting algorithms.

- The fitting algorithm permits a multi-resolution approach so whereas standard N3 practice is to select a single resolution at which bias correction occurs, the N4 framework permits a multi-resolution correction where a base resolution is chosen and correction can then occur at multiple resolution levels each resolution being twice the resolution of the previous level.

Specifically, with respect to segmentation, there exists a third advantage with N4 over N3 in that the former permits the specification of a probabilistic mask as opposed to a binary mask. Recent demonstrations suggest improved white matter segmentation produces better gain field estimates using N3 [64]. Thus, when performing 3-tissue segmentation, we may opt to use, for instance, the posterior probability map of white matter at the current iteration as a weighted mask for input to N4. This is done by setting the '--weight-image' option on the N4 command line call (see Listing 2) to the posterior probability image corresponding to the white matter produced as output in the Atropos call, i.e. 'Atropos --output'. N4 was recently added to the Insight Toolkit repository[7]

---

[7]http://www.itk.org/Doxygen/html/classitk_1_1N4MRIBiasFieldCorrectionImageFilter.html

where it is built and tested on multiple platforms nightly. The evaluation section will illustrate inclusion of Atropos, N4 and ANTs in a brain processing pipeline.

3.5 Running the Atropos Optimization

The Atropos algorithm is cross-platform and compiles on, at minimum, modern OSX, Windows and Linux-based operating systems. The user interface may be reached through the operating system's user terminal. Because of its portability and low-level efficiency, Atropos can easily be called from within other packages, such as Matlab or Slicer, or, alternatively, integrated at compile time as a library. A typical call to the algorithm, illustrated here with ANTs example data, is: `Atropos -d 2 -a r16slice.nii.gz -i kmeans[3] -c [5,0] -x mask.nii.gz -m [0.2,1x1] -o [r16_seg.nii.gz,r16_prob_%02d.nii.gz]`. In this case, Atropos will output the segmentation image, the per-class probability images and a listing of the parameters used to set up the algorithm. A useful feature is that one may re-initialize the Atropos EM via the `-i PriorProbabilityImages[...]` option. This feature allows one to compute an initial segmentation via K-means, alter the output probabilities by externally computed functions (e.g. Gaussian smoothing, image similarity or edge maps) and re-estimate the segmentation with the modified priors. Finally, the functionality that is available to parametric models is equally available to the non-parametric models enabled by Atropos.

4 Evaluation

Atropos encodes a family of segmentation techniques that may be instantiated for different applications but here we evaluate only two of the many possibilities. First, we perform an evaluation on the BrainWeb dataset using both the standard T1 image with multiple bias and noise levels and also the BrainWeb20 data [65, 66]. In combination, these data allow one to vary not only noise and bias but also the underlying anatomy. Second, we evaluate the use of Atropos in improving whole-brain parcellation and exercise its ability to efficiently solve *many-class* expectation maximization problem. We choose this evaluation problem in part to illustrate the flexibility of Atropos and also the benefits of the novel, efficient implementation that allows many-class problems to be solved with low memory usage ($<$2GB for a 69-class model on 1 mm^3 brain data).

4.1 BrainWeb Evaluation

The BrainWeb data is freely available at `http://mouldy.bic.mni.mcgill.ca/BrainWeb/`. We employ both the individual subject data and the BrainWeb20 data in this evaluation.

4.1.1 Single-Subject Evaluation

We use the single-subject data with 3% noise and three levels of bias referred to as 0, 20 and 40% RF inhomogeneity. We study the effect of the MRF prior term and initialization on the Dice overlap between ground truth and the segmentation result for each tissue. We test both K-means and prior label image initialization with MRF $\beta \in \{0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30\}$ at each bias field. We also feed the white matter probability map derived from K-means into N4 to guide the bias correction.[8] Segmentation is then repeated, with the same parameters, but with the N4-corrected image as input. The resulting algorithm is similar to those that fix segmentation parameters while estimating bias and fix bias while estimating segmentation parameters. Thus, with this simple evaluation, we are able to compare the impact of bias on the combination of N4 and Atropos and also the validity of our prior label image initialization. Results of these evaluation scenarios, in terms of Dice overlap, are shown in Figure 4. Because overlap ratios with N4 bias correction approximate those of the zero bias data, we may conclude that simple N4 pre-processing is adequate to correct even the 40% RF bias level. An example of this procedure, using BrainWeb data with 40% RF bias, is in Figure 3. We supply the information necessary to repeat the results in this figure in the script entitled '`atroposBwebRF40FigureExample.sh`' which is available in the ANTs Atropos documentation folder as of SVN commit 711. The script may be easily modified to run the whole evaluation. Figure 3 shows the results of simultaneously using proton density and T1-weighted BrainWeb data to perform the segmentation. This multivariate input data outperforms the univariate T1-weighted data alone.

4.1.2 20-subject Evaluation

The single-subject BrainWeb study in the previous section tested the basic Atropos options and the benefit of N4 for segmentation in the presence of bias. The 20 subject BrainWeb data allows us to use 2-fold cross-validation to test our ability to segment different individuals reliably.

---

[8]A comprehensive evaluation of N4 reported in [51] used the BrainWeb data set to compare performance with the original N3 algorithm [63].
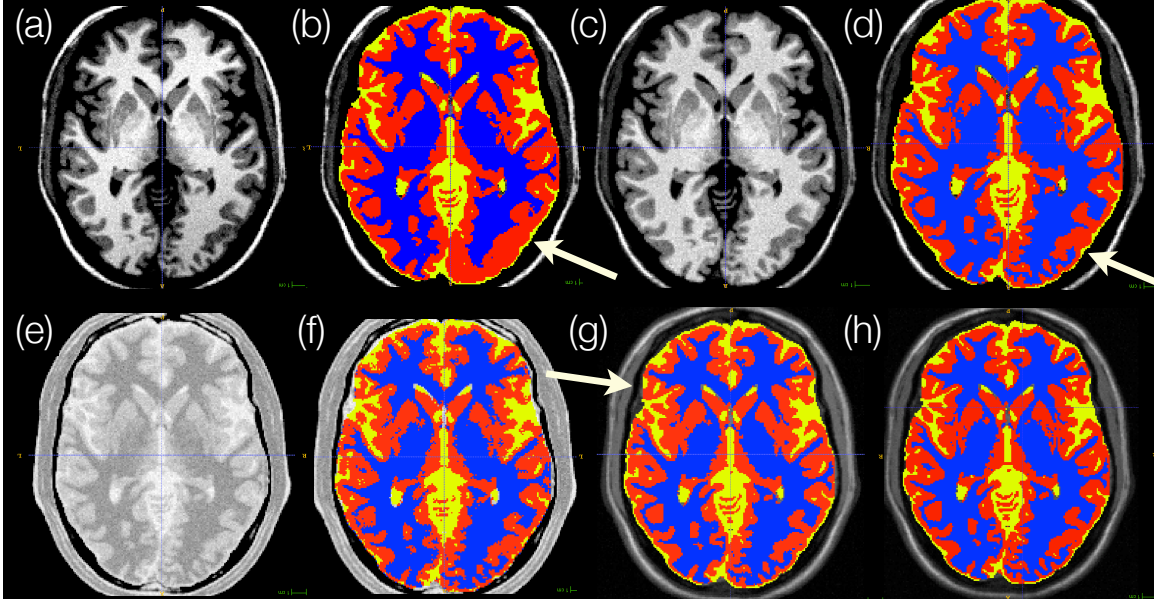
Figure 3: We combine N4 and Atropos by simple sequential processing and apply to BrainWeb T1-weighted single-subject data with 40% RF bias and 3% noise. The β for the MRF term is, here, 0.2. Slice 71 of the input data is in (a). The initial K-means ($K = 3$) segmentation is in panel (b). We use the brain mask to guide N4 bias correction and produce the image in (c). We repeat the K-means segmentation, but with the N4-corrected image as input and produce the segmentation in (d). The average 3-tissue Dice overlap of result (b) is 0.906 while the average overlap for (d) is 0.954. Arrows highlight a region of large before-after segmentation discrepancy. In (e) we see the BrainWeb proton density image with no inhomogeneity and 3% noise. Its segmentation is in (f) with average 3-tissue Dice overlap of 0.895. In (g) we use both proton density data and T1 data as multiple modality input to Atropos. The segmentation of this two-modality input data, using a multivariate Gaussian model, produces average 3-tissue Dice overlap of 0.958, which exceeds the univariate solution. An arrow highlights one region where there is small, visually recognizable improvement in sulcal segmentation relative to the result from T1 data alone. A second area of improvement is the putamen segmentation. The ground truth segmentation is in (h). The multivariate segmentation result, in combination with the low PD segmentation performance, suggests PD and T1 provide complementary information that may improve 3-tissue segmentation and serves to validate the multivariate Atropos implementation. In this case, the benefit is likely to derive from the fact that the PD image has no bias.
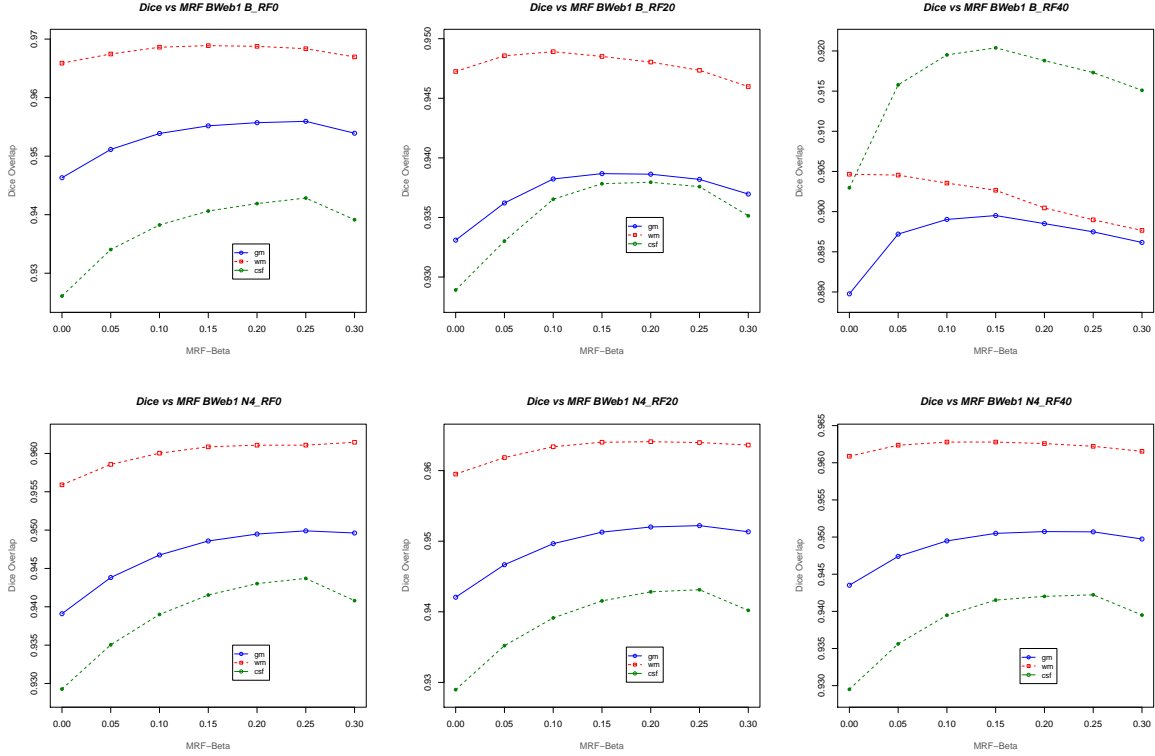
Figure 4: BrainWeb single-subject results for each tissue. The results show that N4 bias correction, combined with Atropos, results in a minimal effect of bias, even at the 40% level. The optimal β for the MRF term appears to be between 0.1 and 0.2. The legend is in the same position in each graph, allowing a visual comparison of the results. As one may see, the N4-assisted overlap values are consistent across bias field/RF inhomogeneity.
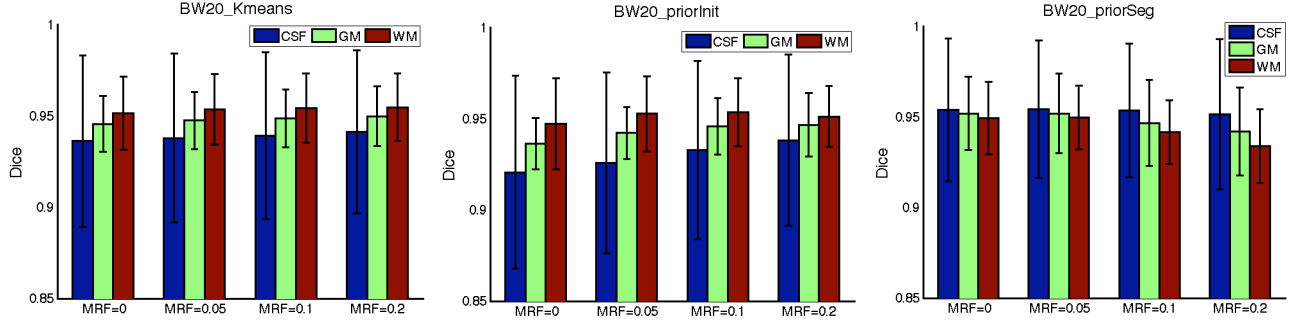
Figure 5: BrainWeb 20-subject results for each tissue as a function of MRF-β parameter where MRF-β is in $\{0, 0.05, 0.1, 0.15, 0.2\}$ and increases left to right. The results show that the PriorProbabilityMaps with $w = 0.5$ (far right) gives the best performance for all tissues.

In this study, we divide the 20 subjects equally into training and testing groups. We then exploit the ground-truth labeling of the training data to build both a group template [49] and also prior probability maps for each of the three major tissues in the cerebrum. Each prior probability map is gained by deforming the ground truth labels from each of the 10 training subjects to their template and averaging component by component. We then deform the template—and priors—to the ten testing subjects and run Atropos with not only `KMeans[3]` initialization but also `PriorProbabilityMap[3,priors%02d.nii.gz,`w`]` where $w \in \{0.0, 0.5\}$. We then switch the roles of testing and training sets to gain 3-tissue segmentation for each of the twenty subjects. When $w = 0$, the priors are only used in initializing the model parameters but not during subsequent EM iterations. When $w = 0.5$, the priors are maintained in the product with the likelihood during all EM iterations. Results, in terms of bar plots for Dice overlap mean and standard deviation, are shown in Figure 5.

4.2   The Hammers Dataset Evaluation

We evaluate the ability to improve multi-template labeling results by converting the group labels to probability maps and using them to drive many-class EM segmentation. The ground truth labels cover 69 classes and much of the brain. Some unlabeled regions remain which we assign to label 69 such that all brain parenchyma contains a unique label. Following [50], the initialization of our evaluation applies the script `ants_multitemplate_labeling.sh` (available in ANTs) to the 19 Hammers evaluation datasets located at `http://www.brain-development.org/` and currently under the adult atlases section [38, 67]. These initial majority voting results are competitive with

prior work [38, 68] and serve as a baseline against which we compare.

We first convert each of the 69 labels within the original evaluation dataset to an individual image. The remaining steps, summarized briefly, are the same for each of the 19 subjects. We select one subject as an unlabeled target. The other 18 datasets are then mapped (as in the script above) to that subject. We then deform, individually, the 69×18 label images to the unlabeled subject. The label probability map is gained by averaging the 18 deformed images associated with each label. We repeat this for each subject. The following parameters are the most relevant to this discussion: `-i PriorProbabilityImages[3,label_prob%02d.nii.gz,0.5] -m [0.1,1x1x1] -c [5,0] -p Socrates[1]`. Results, in terms of Dice overlap, are shown in Figure 6.

4.3 Reproducibility of this Evaluation

Data: The BrainWeb data is freely available. We used single-subject BrainWeb data as is but added a metaformat data header to the raw binary files. An example copy of this header is contained in `atroposBwebRF40FigureExample.sh` . The 20-subject data, however, required excluding non-cerebrum tissue classes. The Hammers data was also used as is (`http://www.brain-development.org/`).

Software: The ANTs software is available at `http://www.picsl.upenn.edu/ANTs` with download and compilation instructions at `http://www.picsl.upenn.edu/ANTS/download.php`. SVN release 711 was used for the examples and evaluations performed in this paper. Some components of ANTs depend on the Insight ToolKit. The most critical dependency, for Atropos, is the ITK statistics framework used to implement the univariate and multivariate parametric models. We linked to the git version of ITK current as of Dec. 1, 2010. See `http://www.itk.org/Wiki/ITK/Git/Download` for instructions on git ITK.

Scripts: The complete script for the single-subject BrainWeb study is based on generalizing `atroposBwebRF40FigureExample.sh`, which is available in the ANTs toolkit (SVN release 711 or greater) and which reproduces Figure 2 results. The template-based normalization procedure for the BrainWeb 20 and the Hammers evaluation data is based on freely available scripts included with ANTs, `ants_multitemplate_labeling.sh` and `buildtemplateparallel.sh`. A release version of ANTs—with a final version of Atropos—will be prepared with the final version of this manuscript.

Neuroanatomical Region	Label	Atropos	Majority	PVal	Sign
Right_Hippocampus	1	0.8522	0.8347	0.000183	+
Left_Hippocampus	2	0.8363	0.8234	0.005046	+
Right_Amygdala	3	0.832	0.8029	5.53E-08	+
Left_Amygdala	4	0.8346	0.8108	2.16E-05	+
Right_Anterior_temporal_lobe_medial_part	5	0.9078	0.8822	1.41E-12	+
Left_Anterior_temporal_lobe_medial_part	6	0.9092	0.8838	6.04E-12	+
Right_Anterior_temporal_lobe_lateral_part	7	0.8935	0.8652	7.55E-15	+
Left_Anterior_temporal_lobe_lateral_part	8	0.9022	0.8742	1.46E-13	+
Right_Parahippocampal_and_ambient_gyri	9	0.8651	0.8421	1.90E-10	+
Left_Parahippocampal_and_ambient_gyri	10	0.866	0.8441	3.91E-10	+
Right_Superior_temporal_gyrus	11	0.8907	0.874	1.16E-11	+
Left_Superior_temporal_gyrus	12	0.8976	0.8798	6.76E-09	+
Right_Middle_and_inferior_temporal_gyri	13	0.89	0.874	1.27E-09	+
Left_Middle_and_inferior_temporal_gyri	14	0.8908	0.873	9.39E-11	+
Right_Fusiform_gyrus	15	0.7667	0.7461	6.79E-08	+
Left_Fusiform_gyrus	16	0.7556	0.7391	0.000685	+
Right_Cerebellum	17	0.9829	0.9702	6.94E-18	+
Left_Cerebellum	18	0.9833	0.9703	3.72E-17	+
Brainstem	19	0.9523	0.9544	0.159323	~
Left_Insula	20	0.8812	0.8797	0.06472	~
Right_Insula	21	0.8754	0.8738	0.18553	~
Left_Lateral_remainder_of_occipital_lobe	22	0.8603	0.8444	1.05E-12	+
Right_Lateral_remainder_of_occipital_lobe	23	0.8594	0.8434	3.70E-16	+
Left_Gyrus_cinguli_anterior_part	24	0.7671	0.7842	0.00263	-
Right_Gyrus_cinguli_anterior_part	25	0.8234	0.8407	0.000209	-
Left_Gyrus_cinguli_posterior_part	26	0.8332	0.8418	0.001878	-
Right_Gyrus_cinguli_posterior_part	27	0.8154	0.8233	0.015334	~
Left_Middle_frontal_gyrus	28	0.8674	0.8554	1.50E-11	+
Right_Middle_frontal_gyrus	29	0.8717	0.8608	4.93E-15	+
Left_Posterior_temporal_lobe	30	0.872	0.8617	2.52E-10	+
Right_Posterior_temporal_lobe	31	0.8683	0.8578	9.39E-11	+
Left_Inferolateral_remainder_of_parietal_lobe	32	0.8706	0.8581	1.98E-11	+
Right_Inferolateral_remainder_of_parietal_lo	33	0.8573	0.8434	2.08E-11	+
Left_Caudate_nucleus	34	0.8829	0.8952	5.81E-06	-
Right_Caudate_nucleus	35	0.8853	0.8921	0.000305	-
Left_Nucleus_accumbens	36	0.7361	0.7382	0.822167	~
Right_Nucleus_accumbens	37	0.7204	0.7158	0.406373	~
Left_Putamen	38	0.8925	0.8951	0.215131	~
Right_Putamen	39	0.8936	0.8981	0.015623	~
Left_Thalamus	40	0.9089	0.9148	0.002933	-
Right_Thalamus	41	0.9054	0.9102	0.003462	-
Left_Pallidum	42	0.7996	0.8288	0.000563	-
Right_Pallidum	43	0.7966	0.8284	0.002293	-
Corpus_callosum	44	0.839	0.886	3.02E-09	-
Right_Lateral_ventricle_excluding_tempora	45	0.8911	0.89	0.581331	~
Left_Lateral_ventricle_excluding_temporal_	46	0.9003	0.8986	0.408928	~
Right_Lateral_ventricle_temporal_horn	47	0.6557	0.625	0.037556	~
Left_Lateral_ventricle_temporal_horn	48	0.6523	0.62	0.001961	+
Third_ventricle	49	0.8313	0.817	9.85E-05	+
Left_Precentral_gyrus	50	0.8297	0.824	0.001547	+
Right_Precentral_gyrus	51	0.8463	0.8365	1.15E-06	+
Left_Gyrus_rectus	52	0.8132	0.7991	8.73E-08	+
Right_Gyrus_rectus	53	0.8245	0.8062	8.53E-08	+
Left_Orbitofrontal_gyri	54	0.8625	0.8451	3.75E-11	+
Right_Orbitofrontal_gyri	55	0.8787	0.8599	1.69E-12	+
Left_Inferior_frontal_gyrus	56	0.8491	0.833	3.51E-11	+
Right_Inferior_frontal_gyrus	57	0.8429	0.8282	1.49E-09	+
Left_Superior_frontal_gyrus	58	0.8806	0.8679	1.42E-10	+
Right_Superior_frontal_gyrus	59	0.8894	0.8778	6.43E-11	+
Left_Postcentral_gyrus	60	0.8119	0.7949	1.74E-06	+
Right_Postcentral_gyrus	61	0.8189	0.8009	7.95E-08	+
Left_Superior_parietal_gyrus	62	0.8604	0.8472	6.32E-10	+
Right_Superior_parietal_gyrus	63	0.8714	0.8562	7.95E-08	+
Left_Lingual_gyrus	64	0.855	0.8371	1.87E-08	+
Right_Lingual_gyrus	65	0.8453	0.8259	6.71E-08	+
Left_Cuneus	66	0.834	0.8077	4.66E-10	+
Right_Cuneus	67	0.8515	0.8278	5.93E-08	+
Unlabeled	68	0.7767	0.7579	0.002256	+
	average	0.851	0.8412		

Figure 6: The figure compares the Dice overlap results from Atropos versus the raw results from majority voting for each of 68 neuroanatomical regions and, in addition, the unlabeled portions of the brain from the Hammers evaluation dataset. We evaluated Atropos via N-fold cross-validation and employed PriorProbabilityImages for each class where probabilities are gained by averaging mapped subject labels. The color coding highlights those regions that have the highest (yellow) and lowest (pink) improvement. The significance of the improvement, measured by pairwise T-test, is also shown as is a trinary coding of that improvement as: + significant improvement, − performance reduction, ∼ no change.

5   Discussion

We introduced Atropos, the theory and implementation details and documented its performance in a variety of use cases. We also showed evidence that the openly available N4 bias correction can easily be used with Atropos to improve segmentation. Furthermore, we used multiple subject BrainWeb data to build dataset-specific priors that provided the most consistent segmentation performance across tissues. Finally, we used majority voting to initialize an Atropos EM solution to a 69-class brain parcellation problem. Significant improvements were gained in multiple brain regions, in particular in temporal lobe cortex, the hippocampi and amygdalae and the lateral ventricles. This work, in summary, proves the applicability of Atropos in both basic and extended use cases.

5.1   Performance on BrainWeb Data

Atropos results are competitive with the state of the art. For instance, [28] (SPM5) evaluated on 0% RF bias field, 3% noise BrainWeb single subject data finding 0.932 (GM) and .961 (WM) Dice overlap. Results on 40% RF bias were 0.934 (GM) and 0.961 (WM). SPM5 exhibits insensitivity to bias similar to our own best results on the 40% RF bias, 3% noise case (MRF-β=0.2, K-means + N4) with Dice overlap for GM is 0.951 and for WM is 0.963. [69] gave GM Dice overlap results (BrainWeb single 3% noise) of 0.962 (0% RF bias), 0.964 (20% RF bias) and 0.956 (40% RF bias) which are slightly higher than either SPM5 or Atropos results. However, [69] do not report WM or CSF numbers for comparison. Topology-preserving methods also perform well. [70] achieved Dice overlap for 3% noise 20% RF bias BrainWeb single subject with 0.927, 0.912 and 0.900 for WM, GM and CSF, respectively. These are excellent numbers given the topological constraint applied to the segmentation. [71] proposed TOADS and, estimating from the paper's graph, showed that the average Dice overlap accuracy for 3% noise for various RF was 0.950–0.960 for WM, 0.930–0.950 for GM and 0.920–0.940 for CSF. Perhaps the most recent balanced evaluation was performed in [12], which reports confusion matrix numbers, rather than Dice overlap. Because the absolute true number of GM and WM voxels for BrainWeb are known, we can convert the confusion matrix to Dice overlap. In that case, the SPM5 Dice overlap for BrainWeb single GM and WM is 0.885 and 0.909, respectively, while FreeSurfer and FSL's accuracy is lower. The best GM result for 20 subjects

is obtained by SPM5: 0.930; best WM Dice overlap is from FSL: 0.950. We note that [12] used a comprehensive evaluation where quality of brain extraction also contributed to the outcome. Thus, the results must be interpreted slightly differently than those from other papers. Finally, we note that in our evaluation of 20-subject BrainWeb data, the prior probability models performed best of all the models used. Compared to the K-Means based segmentation, the prior based segmentation performance also peaked at lower values of the MRF-β term (0.0 and 0.05). This is reasonable in that the spatial priors themselves impose a degree of regularity on the segmentation, as in SPM5.

5.2   Performance on Hammers Data

Our prior work, [48], showed that the majority vote initialization provided to Atropos by ANTs template mapping is competitive with [38]. Overall, the Atropos EM extension improved these results further. However, in a few regions of the mid-brain, the Atropos EM segmentation performed significantly worse. This is not surprising, in that Atropos EM assumes that signal from the likelihood and MRF term is valuable in improving the segmentation. This assumption held for amygdala and lateral ventricles among other areas. However, in pallidum and corpus callosum (the most significant areas with loss of performance), this is not true. We believe the explanation is that the intensity varies within these structures and that a more complex intensity model (or finer parcellation) would be needed here. An alternative solution would be to use boundary conditions for these structures, as in the `PriorLabelImage` Atropos initialization option.

5.3   Clinically-Related Evaluation

While specifying performance on BrainWeb is highly valuable, clinical validation is a second important aspect of segmentation evaluation. For instance, [44, 72–75] are only a few of the papers that evaluate segmentation performance with respect to a known neurobiological outcome measure. Atropos is currently used in clinical studies and a number of clinically focused, application-specific evaluations are ongoing and will constitute future work. One early example of a clinically-focused Atropos neuroimaging application is in [76]. A second successful application area is that of ventilation-based segmentation of hyperpolarized helium-3 MRI [77] which also used the open source Glamorous Glue algorithm to impose topology constraints [78]. Thus, future work may incorporate topology more closely into the Atropos methodology.

A more general advantage which extends beyond the scope of the experimental evaluation

section of this paper is the flexibility of Atropos. This includes not only n-tissue segmentation and dense volumetric cortical parcellation, as reported in this work, but Atropos is also used in conjunction with our ANTs registration tools for robust brain extraction which has reported good performance in comparison with other popular, publicly available brain extraction tools [50].

6   Conclusion

The Atropos software is freely available to the public. We release this code not only to make it available to clinical researchers but with the hope that other researchers in segmentation will provide feedback about the implementation decisions that we made. EM segmentation is non-trivial and there are numerous design alternatives available not only in the models selected but also in the ICM coding, alternatives to ICM and the method in which prior and likelihood are combined. Due to the flexibility of Atropos, we also hope that some of its capabilities, though not evaluated here, are explored by the segmentation or clinical community.

Information Sharing Statement   Atropos software is available in ANTs http://www.picsl. upenn.edu/ANTs which depends on ITK http://www.itk.org/Wiki/ITK/Git/Download. The data used in this work is available in the ANTs software repository, BrainWeb http://mouldy. bic.mni.mcgill.ca/BrainWeb/ and at www.brain-development.org. We employed itk-SNAP for visualization www.itksnap.org.

References

[1] Vannier M.W., Butterfield R.L., Jordan D., Murphy W.A., Levitt R.G., and Gado M. (1985) Multispectral analysis of magnetic resonance images. Radiology 154, 221–224.

[2] Bezdek J.C., Hall L.O., and Clarke L.P. (1993) Review of MR image segmentation techniques using pattern recognition. Med. Phys. 20, 1033–1048.

[3] Pal N.R. and Pal S.K. (1993) A review on image segmentation techniques. Pattern Recognition 26, 1277–1294.

[4] Clarke L.P., Velthuizen R.P., Camacho M.A., Heine J.J., Vaidyanathan M., Hall L.O., Thatcher R.W., and Silbiger M.L. (1995) MRI segmentation: methods and applications. Magn. Reson. Imaging 13, 343–368.

[5] Pham D.L., Xu C., and Prince J.L. (2000) Current methods in medical image segmentation. Annu. Rev. Biomed. Eng. 2, 315–337.

[6] Viergever M.A., Maintz J.B., Niessen W.J., Noordmans H.J., Pluim J.P., Stokking R., and Vincken K.L. (2001) Registration, segmentation, and visualization of multimodal brain images. Comput. Med. Imaging. Graph. 25, 147–151.

[7] Suri J.S., Singh S., and Reden L. (2002) Computer vision and pattern recognition techniques for 2-D and 3-D MR cerebral cortical segmentation (part I): A state-of-the-art review. Pattern Analysis & Applications 5, 46–76. 10.1007/s100440200005.

[8] Duncan J.S., Papademetris X., Yang J., Jackowski M., Zeng X., and Staib L.H. (2004) Geometric strategies for neuroanatomic analysis from MRI. Neuroimage 23 Suppl 1, S34–S45.

[9] Balafar M.A., Ramli A.R., Saripan M.I., and Mashohor S. (2010) Review of brain MRI image segmentation methods. Artif. Intell. Rev. 33, 261–274.

[10] Cuadra M.B., Cammoun L., Butz T., Cuisenaire O., and Thiran J.P. (2005) Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. IEEE Trans. Med. Imaging. 24, 1548–1565.

[11] Zaidi H., Ruest T., Schoenahl F., and Montandon M.L. (2006) Comparative assessment of statistical brain MR image segmentation algorithms and their impact on partial volume correction in PET. Neuroimage 32, 1591–1607.

[12] Klauschen F., Goldman A., Barra V., Meyer-Lindenberg A., and Lundervold A. (2009) Evaluation of automated brain MR image segmentation and volumetry methods. Hum. Brain. Mapp. 30, 1310–1327.

[13] de Boer R., Vrooman H.A., Ikram M.A., Vernooij M.W., Breteler M.M.B., van der Lugt A., and Niessen W.J. (2010) Accuracy and reproducibility study of automatic MRI brain tissue segmentation methods. Neuroimage 51, 1047–1056.

[14] Dempster A., Laird N., and Rubin D. (1977) Maximum likelihood estimation from incomplete data using the EM algorithms. J. Royal Stat. Soc. 39, 1–38.

[15] Wells W.M., Grimson W.L., Kikinis R., and Jolesz F.A. (1996) Adaptive segmentation of MRI data. IEEE Trans Med Imaging 15, 429–442.

[16] Cline H.E., Lorensen W.E., Kikinis R., and Jolesz F. (1990) Three-dimensional segmentation of MR images of the head using probability and connectivity. J Comput Assist Tomogr 14, 1037–1045.

[17] Kikinis R., Shenton M.E., Gerig G., Martin J., Anderson M., Metcalf D., Guttmann C.R., McCarley R.W., Lorensen W., and Cline H. (1992) Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. J Magn Reson Imaging 2, 619–629.

[18] Weisenfeld N.I. and Warfield S.K. (2009) Automatic segmentation of newborn brain MRI. Neuroimage 47, 564–572.

[19] Geman S. and Geman D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern. Anal. Mach. Intell. 6, 721–741.

[20] Held K., Kops E.R., Krause B.J., Wells W.M., Kikinis R., and Mller-Grtner H.W. (1997) Markov random field segmentation of brain MR images. IEEE Trans Med Imaging 16, 878–886.

[21] Leemput K.V., Maes F., Vandermeulen D., and Suetens P. (1999) Automated model-based bias field correction of MR images of the brain. IEEE Trans Med Imaging 18, 885–896.

[22] Leemput K.V., Maes F., Vandermeulen D., and Suetens P. (1999) Automated model-based tissue classification of MR images of the brain. IEEE Trans Med Imaging 18, 897–908.

[23] Zhang Y., Brady M., and Smith S. (2001) Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20, 45–57.

[24] Scherrer B., Forbes F., Garbay C., and Dojat M. (2009) Distributed local MRF models for tissue and structure brain segmentation. IEEE Trans Med Imaging 28, 1278–1295.

[25] Pappas T.N. (1992) An adaptive clustering algorithm for image segmentation. IEEE Trans. Signal Proc. 40, 901–914.

[26] Greenspan H., Ruf A., and Goldberger J. (2006) Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. IEEE Trans Med Imaging 25, 1233–1245.

[27] Marroquin J.L., Vemuri B.C., Botello S., Calderon F., and Fernandez-Bouzas A. (2002) An accurate and efficient Bayesian method for automatic segmentation of brain MRI. IEEE Trans Med Imaging 21, 934–945.

[28] Ashburner J. and Friston K.J. (2005) Unified segmentation. Neuroimage 26, 839–851.

[29] Ruan S., Jaggi C., Xue J., Fadili J., and Bloyet D. (2000) Brain tissue classification of magnetic resonance images using partial volume modeling. IEEE Trans Med Imaging 19, 1179–1187.

[30] Ballester M.A.G., Zisserman A.P., and Brady M. (2002) Estimation of the partial volume effect in MRI. Med Image Anal 6, 389–405.

[31] Leemput K.V., Maes F., Vandermeulen D., and Suetens P. (2003) A unifying framework for partial volume segmentation of brain MR images. IEEE Trans Med Imaging 22, 105–119.

[32] Friston K.J., Frith C.D., Liddle P.F., Dolan R.J., Lammertsma A.A., and Frackowiak R.S. (1990) The relationship between global and local changes in PET scans. J Cereb Blood Flow Metab 10, 458–466.

[33] Pohl K.M., Fisher J., Grimson W.E.L., Kikinis R., and Wells W.M. (2006) A Bayesian model for joint segmentation and registration. Neuroimage 31, 228–239.

[34] Pohl K.M., Bouix S., Nakamura M., Rohlfing T., McCarley R.W., Kikinis R., Grimson W.E.L., Shenton M.E., and Wells W.M. (2007) A hierarchical algorithm for MR brain image parcellation. IEEE Trans Med Imaging 26, 1201–1212.

[35] Pieper S., Lorensen B., Schroeder W., and Kikinis R. (2006) The NA-MIC kit: ITK, VTK, pipelines, grids and 3D Slicer as an open platform for the medical image computing community. In Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro. vol. 1, 698–701.

[36] Goualher G.L., Procyk E., Collins D.L., Venugopal R., Barillot C., and Evans A.C. (1999) Automated extraction and variability analysis of sulcal neuroanatomy. IEEE Trans Med Imaging 18, 206–217.

[37] Fischl B., van der Kouwe A., Destrieux C., Halgren E., Sgonne F., Salat D.H., Busa E., Seidman L.J., Goldstein J., Kennedy D., Caviness V., Makris N., Rosen B., and Dale A.M. (2004) Automatically parcellating the human cerebral cortex. Cereb Cortex 14, 11–22.

[38] Heckemann R.A., Hajnal J.V., Aljabar P., Rueckert D., and Hammers A. (2006) Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33, 115–126.

[39] Destrieux C., Fischl B., Dale A., and Halgren E. (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15.

[40] Dale A.M., Fischl B., and Sereno M.I. (1999) Cortical surface-based analysis. I. segmentation and surface reconstruction. Neuroimage 9, 179–194.

[41] Fischl B., Sereno M.I., and Dale A.M. (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207.

[42] Klein A. and Hirsch J. (2005) Mindboggle: a scatterbrained approach to automate brain labeling. Neuroimage 24, 261–280.

[43] Smith S.M., Rao A., Stefano N.D., Jenkinson M., Schott J.M., Matthews P.M., and Fox N.C. (2007) Longitudinal and cross-sectional analysis of atrophy in alzheimer's disease: cross-validation of BSI, SIENA and SIENAX. Neuroimage 36, 1200–1206.

[44] de Bresser J., Portegies M.P., Leemans A., Biessels G.J., Kappelle L.J., and Viergever M.A. (2011) A comparison of MR based segmentation methods for measuring brain atrophy progression. Neuroimage 54, 760–768.

[45] Wolpert D.H. and Macready W.G. (1997) No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation 1, 67–82.

[46] Prastawa M., Gilmore J.H., Lin W., and Gerig G. (2005) Automatic segmentation of MR images of the developing newborn brain. Med Image Anal 9, 457–466.

[47] Boykov Y. and Kolmogorov V. (2004) An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans Pattern Anal Mach Intell 26, 1124–1137.

[48] Avants B.B., Tustison N.J., Song G., Cook P.A., Klein A., and Gee J.C. (2011) A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54, 2033–2044.

[49] Avants B.B., Yushkevich P., Pluta J., Minkoff D., Korczykowski M., Detre J., and Gee J.C. (2010) The optimal template effect in hippocampus studies of diseased populations. Neuroimage 49, 2457–2466.

[50] Avants B., Klein A., Tustison N., Woo J., , and Gee J.C. (2010) Evaluation of open-access, automated brain extraction methods on multi-site multi-disorder data. In 16th Annual Meeting for the Organization of Human Brain Mapping.

[51] Tustison N.J., Avants B.B., Cook P.A., Zheng Y., Egan A., Yushkevich P.A., and Gee J.C. (2010) N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 29, 1310–1320.

[52] Sanjay-Gopal S. and Hebert T.J. (1998) Bayesian pixel classification using spatially variant finite mixtures and the generalized em algorithm. IEEE Trans Image Process 7, 1014–1028.

[53] Li S.Z. (2001) Markov random field modeling in computer vision. Springer-Verlag, London, UK.

[54] Besag J. (1974) Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Royal Statistical Society B-36, 192–236.

[55] Noe A. and Gee J.C. (2001) Partial volume segmentation of cerebral MRI scans with mixture model clustering. In M. Insana and R. Leahy, eds., Information Processing in Medical Imaging, Springer Berlin / Heidelberg, vol. 2082 of *Lecture Notes in Computer Science*, 423–430.

[56] Lim K.O. and Pfefferbaum A. (1989) Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter. J Comput Assist Tomogr 13, 588–593.

[57] Julin P., Melin T., Andersen C., Isberg B., Svensson L., and Wahlund L.O. (1997) Reliability of interactive three-dimensional brain volumetry using MP-RAGE magnetic resonance imaging. Psychiatry Res 76, 41–49.

[58] Freeborough P.A., Fox N.C., and Kitney R.I. (1997) Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. Comput Methods Programs Biomed 53, 15–25.

[59] Yushkevich P.A., Piven J., Hazlett H.C., Smith R.G., Ho S., Gee J.C., and Gerig G. (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31, 1116–1128.

[60] Boykov Y.Y. and Jolly M.P. (2001) Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In Proc. Eighth IEEE Int. Conf. Computer Vision ICCV 2001. vol. 1, 105–112.

[61] Besag J. (1986) On the statistical analysis of dirty pictures. Journal of the Royal Royal Statistical Society, Series B 48, 259–302.

[62] Awate S.P., Tasdizen T., Foster N., and Whitaker R.T. (2006) Adaptive Markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification. Med Image Anal 10, 726–739.

[63] Sled J.G., Zijdenbos A.P., and Evans A.C. (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17, 87–97.

[64] Boyes R.G., Gunter J.L., Frost C., Janke A.L., Yeatman T., Hill D.L.G., Bernstein M.A., Thompson P.M., Weiner M.W., Schuff N., Alexander G.E., Killiany R.J., DeCarli C., Jack C.R., Fox N.C., and Study A.D.N.I. (2008) Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. Neuroimage 39, 1752–1762.

[65] Aubert-Broche B., Griffin M., Pike G.B., Evans A.C., and Collins D.L. (2006) Twenty new digital brain phantoms for creation of validation image data bases. IEEE Trans Med Imaging 25, 1410–1416.

[66] Battaglini M., Smith S.M., Brogi S., and Stefano N.D. (2008) Enhanced brain extraction improves the accuracy of brain atrophy estimation. Neuroimage 40, 583–589.

[67] Hammers A., Allom R., Koepp M.J., Free S.L., Myers R., Lemieux L., Mitchell T.N., Brooks D.J., and Duncan J.S. (2003) Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum Brain Mapp 19, 224–247.

[68] Heckemann R.A., Keihaninejad S., Aljabar P., Rueckert D., Hajnal J.V., Hammers A., and Initiative A.D.N. (2010) Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. Neuroimage 51, 221–227.

[69] Nakamura K. and Fisher E. (2009) Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients. Neuroimage 44, 769–776.

[70] Shiee N., Bazin P.L., Ozturk A., Reich D.S., Calabresi P.A., and Pham D.L. (2010) A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. Neuroimage 49, 1524–1535.

[71] Bazin P.L. and Pham D.L. (2007) Topology-preserving tissue classification of magnetic resonance brain images. IEEE Trans Med Imaging 26, 487–496.

[72] Freeborough P.A. and Fox N.C. (1997) The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Trans Med Imaging 16, 623–629.

[73] Westlye L.T., Walhovd K.B., Dale A.M., Espeseth T., Reinvang I., Raz N., Agartz I., Greve D.N., Fischl B., and Fjell A.M. (2009) Increased sensitivity to effects of normal aging and Alzheimer's disease on cortical thickness by adjustment for local variability in gray/white contrast: a multi-sample MRI study. Neuroimage 47, 1545–1557.

[74] Sánchez-Benavides G., Gmez-Ansn B., Sainz A., Vives Y., Delfino M., and Pea-Casanova J. (2010) Manual validation of Freesurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer disease subjects. Psychiatry Res 181, 219–225.

[75] Chou Y.Y., Lepor N., Avedissian C., Madsen S.K., Parikshak N., Hua X., Shaw L.M., Trojanowski J.Q., Weiner M.W., Toga A.W., Thompson P.M., and Initiative A.D.N. (2009) Mapping correlations between ventricular expansion and CSF amyloid and tau biomarkers in 240 subjects with Alzheimer's disease, mild cognitive impairment and elderly controls. Neuroimage 46, 394–410.

[76] Avants B., Cook P.A., McMillan C., Grossman M., Tustison N.J., Zheng Y., and Gee J.C. (2010) Sparse unbiased analysis of anatomical variance in longitudinal imaging. Med Image Comput Comput Assist Interv 13, 324–331.

[77] Tustison N., Avants B., Altes T., de Lange E., Mugler J., and Gee J. (2010) Automatic segmentation of ventilation defects in hyperpolarized 3He MRI. In Proceedings of the Biomedical Engineering Society Annual Meeting.

[78] Tustison N., Avants B., Siqueira M., and Gee J. (2010) Topological well-composedness and Glamorous Glue: A digital gluing algorithm for topologically constrained front propagation. IEEE Trans Image Process .