

# TreeWAS: Model and Algorithms

Lino A.F. Ferreira

11 July 2020

These notes describe the TreeWAS model and associated inference algorithms introduced by [Cortes et al. \(2017\)](#).

## 1 Model description

Consider a sample of  $N$  individuals indexed by  $i$ . For each individual  $i$ , we observe a series of binary indicators  $z_{ij}$ , with  $j = 1, \dots, J$ , taking the value 1 if this individual has a particular phenotype and 0 otherwise. The indicators  $z_{ij}$  are indexed according to a hierarchical tree structure, with each indicator corresponding to a node in the tree (1, or equivalently  $A$ , is the index of the root node). In our context, the phenotypic indicators represent an ontology of diseases, with increasingly specific disease categories as we move down the tree. We will model these disease indicators  $z_{ij}$  as being the realised values of binary random variables  $Z_{ij}$  and denote this output process by  $\{\mathbf{Z}_i\}_{i=1}^N$ , with  $\mathbf{Z}_i = (Z_{iA}, \dots, Z_{iJ})$ .

We also consider a number of different genetic loci indexed by  $s$ , with  $s = 1, \dots, S$ . At a particular locus  $s$ , the categorical indicator  $G_{is}$  takes the value 0 if individual  $i$  is homozygous for the reference allele at that locus, the value 1 if they are heterozygous and the value 2 if they are homozygous for the alternative allele.

For a *terminal (or leaf) node*  $l$  in the tree, the probability that the random variable  $Z_{il}$  takes value 1 given data  $G_{is}$  is assumed to follow a logistic model:

$$\Pr(Z_{il} = 1 | Y_{ils}) = \frac{e^{Y_{ils}}}{1 + e^{Y_{ils}}}, \quad (1)$$

where  $\Pr(\cdot)$  is the probability mass function (p.m.f.).<sup>1</sup>  $Y_{ils}$  is defined for a subject  $i$ , a terminal node  $l$  and a variant  $s$  as follows:

$$Y_{ils} = \beta_{ls}^0 + \beta_{ls}^1 \mathbb{I}(G_{is} = 1) + \beta_{ls}^2 \mathbb{I}(G_{is} = 2),$$

---

<sup>1</sup>In a slight abuse of notation,  $\Pr(\cdot)$  may denote either a p.m.f. or a probability density function (p.d.f.) depending on the context.

where  $\mathbb{I}(\cdot)$  is the indicator function,  $\beta_{ls}^0$  is the intercept term and  $\beta_{ls}^1$  and  $\beta_{ls}^2$  are separate coefficients for the heterozygous and homozygous states, respectively. From now on, we will consider a single generic locus  $s$  and drop the subscript to simplify notation.

We associate with each node  $j$  in the tree a length-3 vector of coefficients  $(\beta_j^0, \beta_j^1, \beta_j^2)$  and model the pairs  $(\beta_j^1, \beta_j^2)$  as evolving down a tree with the same structure as that of the output process. Given parameter vector  $(\theta, \pi_1)$  the coefficients of a parent node are inherited with probability  $e^{-\theta}$  and transition to a new pair of values with probability  $1 - e^{-\theta}$ . In the latter case, the new values are  $(0, 0)$  with probability  $1 - \pi_1$  while with probability  $\pi_1$  they are drawn from a joint prior  $f(\beta_j^1, \beta_j^2)$ . Writing  $\boldsymbol{\beta}_j = (\beta_j^1, \beta_j^2)$ ,  $f(\boldsymbol{\beta}_j)$  is assumed to be the following:

$$f(\boldsymbol{\beta}_j) \propto \mathcal{N}_2(\mathbf{0}, \Sigma) \cdot \|(\beta_j^1, \beta_j^2/2)\|^k \cdot \epsilon, \quad (2)$$

where  $\mathcal{N}_2$  is a bivariate Normal density with mean vector  $\mathbf{0} = (0, 0)$  and covariance matrix  $\Sigma$ ,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

for some  $\sigma_1 \geq 0$ ,  $\sigma_2 \geq 0$  and  $r \in \mathbb{R}$ , and

$$\epsilon = \begin{cases} 0.1, & \text{if } \beta_j^1 \beta_j^2 < 0 \\ 0.1, & \text{if } |\beta_j^1| > |\beta_j^2| \\ 1, & \text{otherwise.} \end{cases}$$

Note that the p.d.f. in equation (2) is only specified up to a constant of proportionality. The mixture distribution for  $\boldsymbol{\beta}_j$  in the case where a child node does not inherit its parents' coefficients is denoted by  $f^*(\beta_j^1, \beta_j^2)$ . This is also the distribution from which  $\boldsymbol{\beta}_A$ , the vector of coefficients of the root node, is drawn. Cortes et al. (2017) use the following parameter values:  $\pi_1 = 0.001$ ,  $\theta = 1/3$ ,  $\sigma_1 = 2$ ,  $\sigma_2 = 4$ ,  $k = 0.5$  and  $r = 0.5$ .

We make no explicit assumption on the evolution of the  $\beta_j^0$  coefficients. When fitting the model as we describe in the next section, the intercept term will be chosen, for each node  $j$ , to maximise the likelihood of the observed data at that node,  $\mathbf{z}_j = (z_{1j}, \dots, z_{Nj})$ . Figure 1 shows an example of the TreeWAS graphical model.

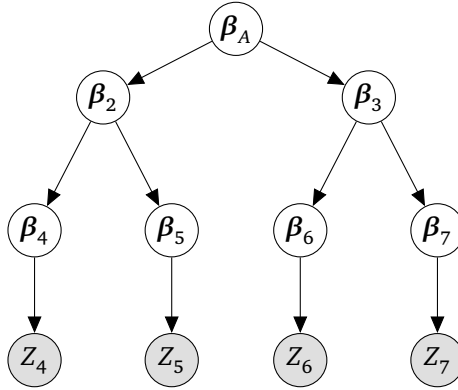


Figure 1: Example tree of disease coefficient pairs  $\beta$  and terminal disease indicator nodes  $Z$ . Only the realised values of the disease indicators (shaded in grey) are observed.

Crucially, note that internal nodes are *not* drawn from a logistic model like the one specified in equation (1). In fact, internal disease nodes are not included in the probabilistic graphical model (see Figure 1) and so we do not write their distribution. While it would be natural to model every node in the tree through a logistic model as above (this would give us a Hidden Markov Tree Model (HMTM) (Crouse et al., 1998)) this would not be compatible with the ontological nature of the disease tree.

To see this, note that the value of every internal node in the disease tree is uniquely determined by that of its children.<sup>2</sup> By this argument, we can say that, conditional on the observed values for its children  $\gamma(j)$ , the value of an internal node  $j$  is fixed (we can write  $z_{ij} = \max\{z_{ik} : k \in \gamma(j)\}$  for all  $i$ ) which is not compatible with each node being stochastically determined based only on its  $\beta$  coefficient vector.

The TreeWAS model is thus a simplified version of the HMTM model where only the terminal hidden state nodes have an associated output with emission distribution specified by equation (1) and only data for the annotations associated with these terminal nodes is used to estimate the model.

<sup>2</sup>For example, consider a parent node corresponding to respiratory infection whose children are nodes corresponding to pneumonia, lung abscess and empyema. If an individual has pneumonia, lung abscess or empyema, then they necessarily have a respiratory infection, and therefore the parent node will take value 1. On the other hand, if they do not have any of these three kinds of respiratory infection then the parent node will take value 0.

## 2 Model fitting

In fitting the model to data, our first goal is to assess, for each genetic locus, whether there is evidence of association between variation at that locus and at least one of the phenotypes in the tree. Having established this, we can then concentrate on those loci for which such evidence of association exists and examine the posterior distributions of the different  $\beta_j$  in their trees.

The first task is one of model comparison between a model under which no evidence of association exists (i.e.  $\beta_j = (0, 0)$  for all nodes  $j$  in the tree) and an alternative model under which there is evidence of association with *at least one* phenotype. The latter model encompasses all instances where at least one coefficient differs from 0.

In keeping with our Bayesian framework, we will use a Bayes Factor to compare these two models. This requires us to compute the integrated likelihood (also known as marginal likelihood or model evidence) of our data under the two scenarios. Defining, for simplicity of notation, a vector  $\beta := (\beta_A, \dots, \beta_J)$  and denoting the terminal output data  $\{\mathbf{z}_l\}_{l=1}^N$  by  $\mathbf{z}$ ,<sup>3</sup> we write the integrated likelihood for the model of no association, denoted  $L_\emptyset$ , as follows:

$$L_\emptyset = \int \Pr(\mathbf{z}, \beta | \beta = \mathbf{0}) d\beta,$$

where  $\Pr(\cdot)$  is now a p.d.f. and  $\mathbf{0}$  is a zero vector of length  $2J$ . Since there is only one way in which all coefficients can be zero, this can be simplified:

$$L_\emptyset = \Pr(\mathbf{z} | \beta = \mathbf{0}).$$

In contrast with the model of no association, there are many (in fact, an infinite number of) ways in which at least one of the model coefficients can differ from zero. Because of this, directly computing the integrated likelihood under the alternative model (denoted  $L_+$ ) would be intractable. To avoid this issue, we note that the full integrated likelihood  $L_{\text{full}}$ , which is defined as

$$L_{\text{full}} = \int \Pr(\mathbf{z}, \beta) d\beta, \tag{3}$$

---

<sup>3</sup>Note that only the most specific disease categories are included in  $\mathbf{z}$ : none of the internal nodes  $z_j$  are included in this model.

can be written as the weighted sum of the null model and alternative model likelihoods:

$$L_{\text{full}} = \pi_{\emptyset} L_{\emptyset} + (1 - \pi_{\emptyset}) L_{+}, \quad (4)$$

where  $\pi_{\emptyset}$  is the prior probability of all coefficients being zero. For a tree with  $J$  nodes, this prior probability is equal to:

$$\pi_{\emptyset} = (1 - \pi_1) \cdot \left[ e^{-\theta} + (1 - e^{-\theta})(1 - \pi_1) \right]^{J-1}.$$

Equation (4) implies that, if we can compute the full and null model likelihoods, we can easily obtain  $L_{+}$ . We can then compute the Bayes Factor to test the null hypothesis of no association against the alternative hypothesis of at least one association:

$$\text{BF}_{\text{tree}} = \frac{L_{+}}{L_{\emptyset}}.$$

Computing the full integrated likelihood  $L_{\text{full}}$  requires numerically evaluating a high-dimensional integral, an expensive operation. Cortes et al. (2017) do this via a recursive algorithm iterating upwards from the terminal nodes of the tree, which makes the computation much more tractable. A small modification of this algorithm then allows for computing  $L_{\emptyset}$ .

A short description of the upward recursive algorithm is as follows: we start by evaluating the likelihood of observing the phenotypic data  $\bar{\mathbf{z}}_l$  for each terminal node  $l$  on the tree for a grid of  $\boldsymbol{\beta}$  values. Moving up one level, we note that the likelihood of an internal node is equal to the likelihood of observing all of its children, which are independent of one another.<sup>4</sup> We thus compute the parents' likelihoods for a grid of  $\boldsymbol{\beta}$  values by multiplying those of their children, and in doing so must integrate over all the values of the children's  $\boldsymbol{\beta}$  coefficients (these are weighted by the prior  $f^*(\boldsymbol{\beta}_l)$ ). We proceed in this manner until reaching the root node. A detailed derivation of this algorithm is given in Appendix A.

We then turn to the question of how to obtain the posterior distributions of the  $\boldsymbol{\beta}_j$  coefficients for those variants for which some evidence of association is found. This requires a second algorithm which proceeds downwards from the root node towards the terminal nodes of the tree, described in detail in Appendix B.

---

<sup>4</sup>This is because the value of a parent node is uniquely determined by those of its children.

Together, these two algorithms are very similar to the the Upward-Downward Algorithm (Crouse et al., 1998) for HMTMs, which in turn is reminiscent of the Forward-Backward Algorithm used for Hidden Markov Models. The presentation of the TreeWAS algorithms in Appendices A and B follows closely that in Durand et al. (2004), although we aim to keep the notation as consistent as possible with that of Cortes et al. (2017) and account for the different structure of our graphical model.

## References

- Cortes, A., C. A. Dendrou, A. Motyer, L. Jostins, D. Vukcevic, A. Dilthey, P. Donnelly, S. Leslie, L. Fugger, and G. McVean (2017). Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK biobank. *Nature Genetics* 49(9), 1311–1318.
- Crouse, M., R. Nowak, and R. Baraniuk (1998). Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on Signal Processing* 46(4), 886–902.
- Durand, J.-B., P. Gonçalves, and Y. Guedon (2004). Computational methods for hidden markov tree models—an application to wavelet trees. *IEEE Transactions on Signal Processing* 52(9), 2551–2560.

## A Upward recursive algorithm

We begin by establishing some additional notation: following [Durand et al. \(2004\)](#), we define  $\bar{Z}_{ij}$  as the *terminal nodes of the subtree* of disease indicators rooted at node  $j$  for individual  $i$ , so that  $\bar{Z}_{iA}$  is the set of all leaf nodes for that individual.  $\bar{Z}_j$  is then the set of terminal nodes of the subtrees rooted at  $j$  for all individuals, and as before  $\bar{z}_j$  is their realised value, i.e. the observed data.

We aim to compute the full integrated likelihood  $L_{\text{full}}$ :

$$L_{\text{full}} = \int \Pr(\bar{\mathbf{z}}_A, \boldsymbol{\beta}) d\boldsymbol{\beta} = \Pr(\bar{\mathbf{z}}_A).$$

Applying simple rules of probability, this can also be written as

$$\begin{aligned} L_{\text{full}} &= \Pr(\bar{\mathbf{z}}_A) \\ &= \int \Pr(\bar{\mathbf{z}}_A, \boldsymbol{\beta}_A) d\boldsymbol{\beta}_A \\ &= \int \Pr(\bar{\mathbf{z}}_A | \boldsymbol{\beta}_A) f^*(\boldsymbol{\beta}_A) d\boldsymbol{\beta}_A. \end{aligned} \quad (5)$$

Following [Cortes et al. \(2017\)](#), we now make two further definitions which will simplify the notation below. We first define  $F_j(\boldsymbol{\beta}_j)$  as the probability of observing the subtree  $\bar{\mathbf{z}}_j$  given only the coefficients  $\boldsymbol{\beta}_j$  for node  $j$ , that is:

$$F_j(\boldsymbol{\beta}_j) := \Pr(\bar{\mathbf{z}}_j | \boldsymbol{\beta}_j).$$

Secondly, and defining  $\rho(j)$  as the parent node of  $j$ , we define  $G_j(\boldsymbol{\beta}_{\rho(j)})$  as the probability of observing the subtree  $\bar{\mathbf{z}}_j$  given only the coefficients  $\boldsymbol{\beta}_{\rho(j)}$  for the parent of node  $j$ , that is:

$$G_j(\boldsymbol{\beta}_{\rho(j)}) := \Pr(\bar{\mathbf{z}}_j | \boldsymbol{\beta}_{\rho(j)}).$$

This now allows us to simplify equation (5) as follows:

$$L_{\text{full}} = \int F_A(\boldsymbol{\beta}_A) f^*(\boldsymbol{\beta}_A) d\boldsymbol{\beta}_A.$$

The key point is that, instead of doing high-dimensional integration to obtain  $\Pr(\bar{\mathbf{z}}_A)$  as in equation (3), we can perform this integration recursively starting from the terminal nodes of the tree. To see this, we

start by noting that, for a terminal node  $l$ ,  $\bar{z}_l = \mathbf{z}_l$ . Then,  $F_l(\boldsymbol{\beta}_l)$  is simply equal to  $\Pr(\mathbf{z}_l | \boldsymbol{\beta}_l)$ . Moreover, assuming independent observations we have  $\Pr(\mathbf{z}_l | \boldsymbol{\beta}_l) = \prod_{i=1}^N \Pr(z_{il} | \boldsymbol{\beta}_l)$ .

For an internal node  $j$ , and defining  $\gamma(j)$  as the children of  $j$ , we can rewrite  $F_j(\boldsymbol{\beta}_j)$  as follows:

$$\begin{aligned} F_j(\boldsymbol{\beta}_j) &:= \Pr(\bar{\mathbf{z}}_j | \boldsymbol{\beta}_j) \\ &= \prod_{k \in \gamma(j)} \Pr(\bar{\mathbf{z}}_k | \boldsymbol{\beta}_j) \\ &= \prod_{k \in \gamma(j)} G_k(\boldsymbol{\beta}_j). \end{aligned}$$

We now note that  $G_k(\boldsymbol{\beta}_j)$  can also be rewritten:

$$\begin{aligned} G_k(\boldsymbol{\beta}_j) &:= \Pr(\bar{\mathbf{z}}_k | \boldsymbol{\beta}_j) \\ &= \int \Pr(\bar{\mathbf{z}}_k, \boldsymbol{\beta}_k | \boldsymbol{\beta}_j) d\boldsymbol{\beta}_k \\ &= \int \underbrace{\Pr(\bar{\mathbf{z}}_k | \boldsymbol{\beta}_k)}_{=F_k(\boldsymbol{\beta}_k)} \Pr(\boldsymbol{\beta}_k | \boldsymbol{\beta}_j) d\boldsymbol{\beta}_k, \end{aligned} \quad (6)$$

where the last equality follows from the HMTM setup, which implies that the terminal phenotypes  $Z_{ik}$  are independent (for all  $i$ ) of the coefficients of their associated parent coefficients, in this case  $\boldsymbol{\beta}_j$ , when conditioned on their own associated coefficients  $\boldsymbol{\beta}_k$ .

The model setup described in Section 1 also implies the following expression for  $\Pr(\boldsymbol{\beta}_k | \boldsymbol{\beta}_j)$ :

$$\Pr(\boldsymbol{\beta}_k | \boldsymbol{\beta}_j) = e^{-\theta} \delta(\boldsymbol{\beta}_k - \boldsymbol{\beta}_j) + (1 - e^{-\theta}) f^*(\boldsymbol{\beta}_k),$$

where  $\delta(\cdot)$  is the Dirac delta function. Equation (6) then becomes:

$$\begin{aligned} G_k(\boldsymbol{\beta}_j) &= \int F_k(\boldsymbol{\beta}_k) \left[ e^{-\theta} \delta(\boldsymbol{\beta}_k - \boldsymbol{\beta}_j) + (1 - e^{-\theta}) f^*(\boldsymbol{\beta}_k) \right] d\boldsymbol{\beta}_k \\ &= e^{-\theta} \int F_k(\boldsymbol{\beta}_k) \delta(\boldsymbol{\beta}_k - \boldsymbol{\beta}_j) d\boldsymbol{\beta}_k + (1 - e^{-\theta}) \int F_k(\boldsymbol{\beta}_k) f^*(\boldsymbol{\beta}_k) d\boldsymbol{\beta}_k. \end{aligned} \quad (7)$$

Noting that

$$\int F_k(\boldsymbol{\beta}_k) \delta(\boldsymbol{\beta}_k - \boldsymbol{\beta}_j) d\boldsymbol{\beta}_k = F_k(\boldsymbol{\beta}_j) = \Pr(\bar{\mathbf{z}}_k | \boldsymbol{\beta}_j),^5$$



and defining a new function  $L_k$  as

$$L_k := \int F_k(\boldsymbol{\beta}_k) f^*(\boldsymbol{\beta}_k) d\boldsymbol{\beta}_k,$$

equation (7) becomes

$$G_k(\boldsymbol{\beta}_j) = e^{-\theta} F_k(\boldsymbol{\beta}_j) + (1 - e^{-\theta}) L_k.$$

To summarise, these are the key equations of the upward recursive algorithm used to compute the full integrated likelihood  $L_{\text{full}}$ :

$$L_{\text{full}} = \int F_A(\boldsymbol{\beta}_A) f^*(\boldsymbol{\beta}_A) d\boldsymbol{\beta}_A \quad (8)$$

$$F_l(\boldsymbol{\beta}_l) = \Pr(\mathbf{z}_l | \boldsymbol{\beta}_l), \quad \text{for a terminal node } l \quad (9)$$

$$F_j(\boldsymbol{\beta}_j) = \prod_{k \in \mathcal{Y}(j)} G_k(\boldsymbol{\beta}_j), \quad \text{for an internal node } j \quad (10)$$

$$G_k(\boldsymbol{\beta}_j) = e^{-\theta} F_k(\boldsymbol{\beta}_j) + (1 - e^{-\theta}) L_k \quad (11)$$

$$L_k = \int F_k(\boldsymbol{\beta}_k) f^*(\boldsymbol{\beta}_k) d\boldsymbol{\beta}_k. \quad (12)$$

The algorithm is then as follows:

1. Initialise by computing  $F_l(\boldsymbol{\beta}_l)$  for all the terminal nodes in the tree using equation (9). This involves computing  $\Pr(\mathbf{z}_l | \boldsymbol{\beta}_l)$  for each terminal node  $l$  for a grid of values for  $\boldsymbol{\beta}_l = (\beta_l^1, \beta_l^2)$ .

As mentioned above,  $\beta_l^0$  is chosen, for each value of  $\boldsymbol{\beta}_l = (\beta_l^1, \beta_l^2)$ , to maximise the likelihood for our logistic model. Since maximum likelihood estimation of the logistic regression model does not have a closed form solution, this must be done numerically for each pair  $\boldsymbol{\beta}_l = (\beta_l^1, \beta_l^2)$  in the grid.

For each terminal node  $l$ , also compute  $L_l$  using equation (12) and  $G_l(\boldsymbol{\beta}_{\rho(l)})$  for a grid of values for  $\boldsymbol{\beta}_{\rho(l)}$  using equation (11);

2. Move up one level and compute  $F_{\rho(l)}(\boldsymbol{\beta}_{\rho(l)})$  using equation (10).  
Also compute  $L_{\rho(l)}$  and  $G_{\rho(l)}$  for a grid of values as before;
3. Continue recursively in this way until reaching the root node and obtaining  $F_A(\boldsymbol{\beta}_A)$ ;
4. Finally, evaluate equation (8) to obtain  $L_A = L_{\text{full}}$ .

---

<sup>5</sup>This is known as the *sifting property* of the delta function.

Note that, using this algorithm, we integrate only over two variables (the different  $\beta_j$  pairs) at a time. In choosing the  $\beta^0$  coefficients to maximise the likelihood, we avoid having to also integrate over them.

Finally, we compute the likelihood of all leaf disease nodes taking value 0, i.e.  $L_\emptyset = \Pr(\bar{\mathbf{z}} | \beta = \mathbf{0})$ , as follows:

$$L_\emptyset = \prod_{l \in T} \Pr(\mathbf{z}_l | \beta_l = \mathbf{0}).$$

## B Downward recursive algorithm

The upward recursion algorithm described in the previous section allows us to compute the full and null likelihoods and, consequently, the likelihood for the model of at least one association. As described above, this can then be used to obtain a Bayes Factor for assessing, for each genetic locus, the evidence of association between this locus and any of the terminal node phenotypes in the tree. Having identified those loci for which such evidence exists, we proceed to deriving the posterior distributions of the different  $\beta_j$  coefficients for these loci. This in turn allows us to quantify the evidence of association between a variant and specific disease phenotypes.

We now describe the downward recursion algorithm used to obtain the posterior distributions of the model's coefficients. First, for a node  $j$ , the posterior distribution of  $\beta_j$  given the data for all terminal nodes  $\bar{\mathbf{z}}_A$  can be written as follows:

$$\Pr(\beta_j | \bar{\mathbf{z}}_A) = \frac{\Pr(\beta_j, \bar{\mathbf{z}}_A)}{\Pr(\bar{\mathbf{z}}_A)}. \quad (13)$$

Since we now know how to obtain  $\Pr(\bar{\mathbf{z}}_A) = L_A$ , by computing the joint distribution  $\Pr(\beta_j, \bar{\mathbf{z}}_A)$  we can obtain the posterior of  $\beta_j$ .

Again following [Durand et al. \(2004\)](#), if  $Z_{ik}$  is a proper subtree of  $Z_{ij}$ , we define  $\bar{Z}_{ij \setminus k}$  as the leaf indicators of the subtree rooted at node  $j$  for individual  $i$  except for the leaf nodes of the subtree rooted at  $k$  for the same individual. Then,  $\bar{\mathbf{z}}_{A \setminus j}$  is all the observed data except for that of subtree  $\bar{\mathbf{z}}_j$ .

We can then rewrite  $\Pr(\beta_j, \bar{\mathbf{z}}_A)$  as follows:

$$\begin{aligned} \Pr(\beta_j, \bar{\mathbf{z}}_A) &= \Pr(\beta_j, \bar{\mathbf{z}}_j, \bar{\mathbf{z}}_{A \setminus j}) \\ &= \Pr(\bar{\mathbf{z}}_j | \beta_j, \bar{\mathbf{z}}_{A \setminus j}) \Pr(\beta_j, \bar{\mathbf{z}}_{A \setminus j}) \\ &= \underbrace{\Pr(\bar{\mathbf{z}}_j | \beta_j)}_{=F_j(\beta_j)} \Pr(\beta_j, \bar{\mathbf{z}}_{A \setminus j}), \end{aligned} \quad (14)$$

where the last equality follows from the fact that  $\bar{\mathbf{Z}}_j$  is independent of  $\bar{\mathbf{Z}}_{A \setminus j}$  when conditioned on  $\beta_j$ . By defining a further function

$$B_j(\beta_j) := \Pr(\beta_j, \bar{\mathbf{z}}_{A \setminus j}),$$

we can write  $\Pr(\beta_j, \bar{\mathbf{z}}_A)$  simply as

$$\Pr(\beta_j, \bar{\mathbf{z}}_A) = F_j(\beta_j) B_j(\beta_j). \quad (15)$$

We now show how  $B_j(\beta_j)$  can be computed through a downward recursion from the root towards the terminal nodes of the tree. For any node other than the root node  $A$ , we have:

$$\begin{aligned} B_j(\beta_j) &:= \Pr(\beta_j, \bar{\mathbf{z}}_{A \setminus j}) \\ &= \int \Pr(\beta_j, \beta_{\rho(j)}, \bar{\mathbf{z}}_{A \setminus \rho(j)}, \bar{\mathbf{z}}_{\rho(j) \setminus j}) d\beta_{\rho(j)} \\ &= \int \frac{\Pr(\beta_j | \beta_{\rho(j)})}{=q(\beta_j, \beta_{\rho(j)})} \Pr(\beta_{\rho(j)}, \bar{\mathbf{z}}_{A \setminus \rho(j)}, \bar{\mathbf{z}}_{\rho(j) \setminus j}) d\beta_{\rho(j)}, \end{aligned}$$

where we denote the transmission distribution of  $\beta_j$  given its parent  $\beta_{\rho(j)}$  by  $q(\beta_j, \beta_{\rho(j)}) := \Pr(\beta_j | \beta_{\rho(j)})$ . Applying the definition of conditional probability and noting that  $\bar{\mathbf{Z}}_{\rho(j)}$  is independent of  $\bar{\mathbf{Z}}_{A \setminus \rho(j)}$  when conditioned on  $\beta_{\rho(j)}$ , we have:

$$\begin{aligned} B_j(\beta_j) &= \int q(\beta_j, \beta_{\rho(j)}) \Pr(\bar{\mathbf{z}}_{\rho(j) \setminus j} | \beta_{\rho(j)}, \bar{\mathbf{z}}_{A \setminus \rho(j)}) \frac{\Pr(\beta_{\rho(j)}, \bar{\mathbf{z}}_{A \setminus \rho(j)})}{=B_{\rho(j)}(\beta_{\rho(j)})} d\beta_{\rho(j)} \\ &= \int q(\beta_j, \beta_{\rho(j)}) \Pr(\bar{\mathbf{z}}_{\rho(j) \setminus j} | \beta_{\rho(j)}) B_{\rho(j)}(\beta_{\rho(j)}) d\beta_{\rho(j)}. \quad (16) \end{aligned}$$

Since the phenotypic child subtrees of a node are independent of one another and of the parent node when conditioned on the parents' coefficients, we have the following:

$$\begin{aligned} B_j(\beta_j) &= \int q(\beta_j, \beta_{\rho(j)}) \frac{\overbrace{\Pr(\bar{\mathbf{z}}_{\rho(j)} | \beta_{\rho(j)})}^{=F_{\rho(j)}(\beta_{\rho(j)})}}{\underbrace{\Pr(\bar{\mathbf{z}}_j | \beta_{\rho(j)})}_{=G_j(\beta_{\rho(j)})}} B_{\rho(j)}(\beta_{\rho(j)}) d\beta_{\rho(j)} \quad (17) \\ &= \int B_{\rho(j)}(\beta_{\rho(j)}) q(\beta_j, \beta_{\rho(j)}) \frac{F_{\rho(j)}(\beta_{\rho(j)})}{G_j(\beta_{\rho(j)})} d\beta_{\rho(j)}. \end{aligned}$$

Since computing  $B_j(\boldsymbol{\beta}_j)$  requires us to first compute  $B_{\rho(j)}(\boldsymbol{\beta}_{\rho(j)})$ , the  $B_j(\boldsymbol{\beta}_j)$  can be computed in a downward recursion starting from the root node, with  $B_A(\boldsymbol{\beta}_A) = \Pr(\boldsymbol{\beta}_A) = f^*(\boldsymbol{\beta}_A)$ .

Finally, having computed  $B_j(\boldsymbol{\beta}_j)$  for a particular node  $j$ , we can see from equations (13) and (15) that the posterior  $\Pr(\boldsymbol{\beta}_j | \bar{\mathbf{z}}_A)$  can be computed as follows:

$$\Pr(\boldsymbol{\beta}_j | \bar{\mathbf{z}}_A) = \frac{F_j(\boldsymbol{\beta}_j)B_j(\boldsymbol{\beta}_j)}{L_A}.$$