Le modèle ornstein-uhlenbeck pour l'évolution d'un trait quantitatif

$$dX_i(t) = \alpha[\beta_i(t) - X_i(t)]dt + \sigma dB_i(t),$$

Le modèle ornstein-uhlenbeck va un pas plus loin que le modèle de mouvement brownien en intégrant une part de sélection à l'évolution du phénotype.

Dans l'équation différentielle ci-dessus, qui décrit la variation du trait, on voit que dX, la variation, dépend de la distance à l'optimum , du trait considéré oû X donne la valeur du trait, et B la valeur idéale pour le trait. Cette force de sélection est multipliée par la constante alpha.

La deuxième composante est le mouvement brownien, ici dB est une variation aléatoire infinitésimale du trait, qui est multipliée par la constante sigma.

Les constantes alpha et sigma donnent respectivement la force de sélection et la force de dérive qui agissent sur le trait.

S'il permet de modéliser l'évolution de manière plus réaliste, ce modèle est aussi plus complexe que le mouvement brownien.
Pour procéder à des simulations avec ce modèle il faut déterminer les paramètres alpha et sigma les plus vraisemblables pour l'arbre phylogénique et les valeurs de traits quantitatifs qu'on utilise.

Le but de mon travail a été la détermination de ces paramètres.

$$\mathcal{L}(\alpha, \sigma, \theta|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \sigma^{2N} \det \tilde{\mathbf{V}}}} \exp\left[-\frac{(\mathbf{x} - \mathbf{W}\boldsymbol{\theta})' \tilde{\mathbf{V}}^{-1} (\mathbf{x} - \mathbf{W}\boldsymbol{\theta})}{2\sigma^2}\right].$$

Cette fonction donne la vraisemblance des paramètres alpha, sigma et theta (la valeur de l'optimum) en fonction des données utilisées.

$$U(\alpha, \sigma, \theta|\mathbf{x}) = N \log 2\pi\sigma^2 + \log \det \tilde{\mathbf{V}} + \frac{1}{\sigma^2} (\mathbf{x} - \mathbf{W}\boldsymbol{\theta})' \tilde{\mathbf{V}}^{-1} (\mathbf{x} - \mathbf{W}\boldsymbol{\theta}).$$

La fonction U est égale à -log(L), minimiser U revient à maximiser L

$$\hat{\sigma}^2(\alpha, \boldsymbol{\theta}) = \frac{1}{N}[\mathbf{x} - \mathbf{W}(\alpha)\boldsymbol{\theta}]'\tilde{\mathbf{V}}^{-1}(\alpha)[\mathbf{x} - \mathbf{W}(\alpha)\boldsymbol{\theta}],$$

$$\hat{\boldsymbol{\theta}}(\alpha) = (\mathbf{W}'\tilde{\mathbf{V}}^{-1}\mathbf{W})\mathbf{W}'\tilde{\mathbf{V}}^{-1}\mathbf{x}.$$

En calculant les dérivées partielles de U, on trouve ces valeurs de sigma et theta pour lesquelles elles s'annulent.

Pour qu'une fonction de plusieurs variables admette un extremum (minimum ou maximum) en un point, il faut que son gradient y soit nul. Le gradient est un vecteur des dérivées partielles de la fonction, c'est donc seulement aux valeurs de sigma et theta donnés par les formules ci-dessus que la fonction peut avoir un minimum.

$$U(\alpha, \hat{\sigma}(\alpha, \hat{\boldsymbol{\theta}}(\alpha)), \hat{\boldsymbol{\theta}}(\alpha)) = N[1 + \log 2\pi\hat{\sigma}^2(\alpha, \hat{\boldsymbol{\theta}}(\alpha))] + \log\det\tilde{\mathbf{V}}(\alpha).$$

En entrant les formules données plus haut dans la fonction U, on obtient cette nouvelle forme, qui ne dépend que de alpha.

La recherche de la bonne valeur de alpha se fait sur ordinateur, en effet elle demande beaucoup de calcul.

Pour chaque valeur de alpha, il faut calculer les matrices :

$$\tilde{V}_{ij} = \frac{e^{-2\alpha(T-s_{ij})}(1 - e^{-2\alpha s_{ij}})}{2\alpha}.$$

T=temps total depuis la racine de l'arbre
Sij=temps oû les espèces I et j ont divergé

$$W_{i0} = e^{-\alpha T},$$

$$W_{ik} = e^{-\alpha T}\sum_{\gamma=1}^{\kappa(i)} \beta_{ik}^{\gamma}(e^{\alpha t_i^{\gamma}} - e^{\alpha t_i^{\gamma-1}}),$$

dans la formule du bas k(i) est l'époque actuelle , une époque désigne le temps après un évènement de spéciation, jusqu'au prochain, en d'autres termes, une branche d'arbre philogénique .
Bik est là au cas oû l'on admet plusieurs optimums dans l'arbre, pour ne pas compter ceux qu'on traite déja ailleurs.

Avec ces matrices, on peut calculer:

$$\hat{\sigma}^2(\alpha, \boldsymbol{\theta}) = \frac{1}{N}[\mathbf{x} - \mathbf{W}(\alpha)\boldsymbol{\theta}]'\tilde{\mathbf{V}}^{-1}(\alpha)[\mathbf{x} - \mathbf{W}(\alpha)\boldsymbol{\theta}],$$

$$\hat{\boldsymbol{\theta}}(\alpha) = (\mathbf{W}'\tilde{\mathbf{V}}^{-1}\mathbf{W})\mathbf{W}'\tilde{\mathbf{V}}^{-1}\mathbf{x}.$$

qui permettent alors de calculer U.

Un optimisateur peut alors trouver la bonne valeur de alpha.