**Variáveis ambientais**

Os modelos de distribuição de espécies (SDMs) serão construídos com base a diferentes tipos de variáveis ambientais combinadas segundo os requerimentos de cada espécies. As variáveis climáticas consistirão nas **19 variáveis bioclimáticas provenientes da plataforma Chelsa 2.1 com 1 km de resolução (Karger et al., 2017).** Adicionalmente será utilizado a altitude do SRTM (https://srtm.csi.cgiar.org) 1 km de resolução e variáveis geomórficas derivadas, com as variáveis topográficas: elevação, inclinação, aspecto, leste, norte, rugosidade, índice de rugosidade do terreno, índice de posição topográfica, vetor medida de robustez, perfil/curvatura tangencial, derivada parcial de primeira/segunda ordem e 10 classes geomorfológicas de relevo, de dados Geomorpho90m com 250 m de resolução (Amatulli et al., 2018; Amatulli et al. 2020). Para as abelhas que utilizam o solo para nidificação também utilizaremos variáveis edáficas relacionadas ao conteúdo de areia, limo e argila, profundidade do solo e pH. Será utilizado o SoilGrid 2.1 (https://soilgrids.org) como fonte de variáveis edáficas com 250 m de resolução. Todas as variáveis serão levadas a uma resolução espacial de ~5 km com uma extensão que abrange a região neotropical.

**CHELSEA com 5km**

**Bio 1 a 19 +**

**cmi**: climate moisture index (*only current conditions*)

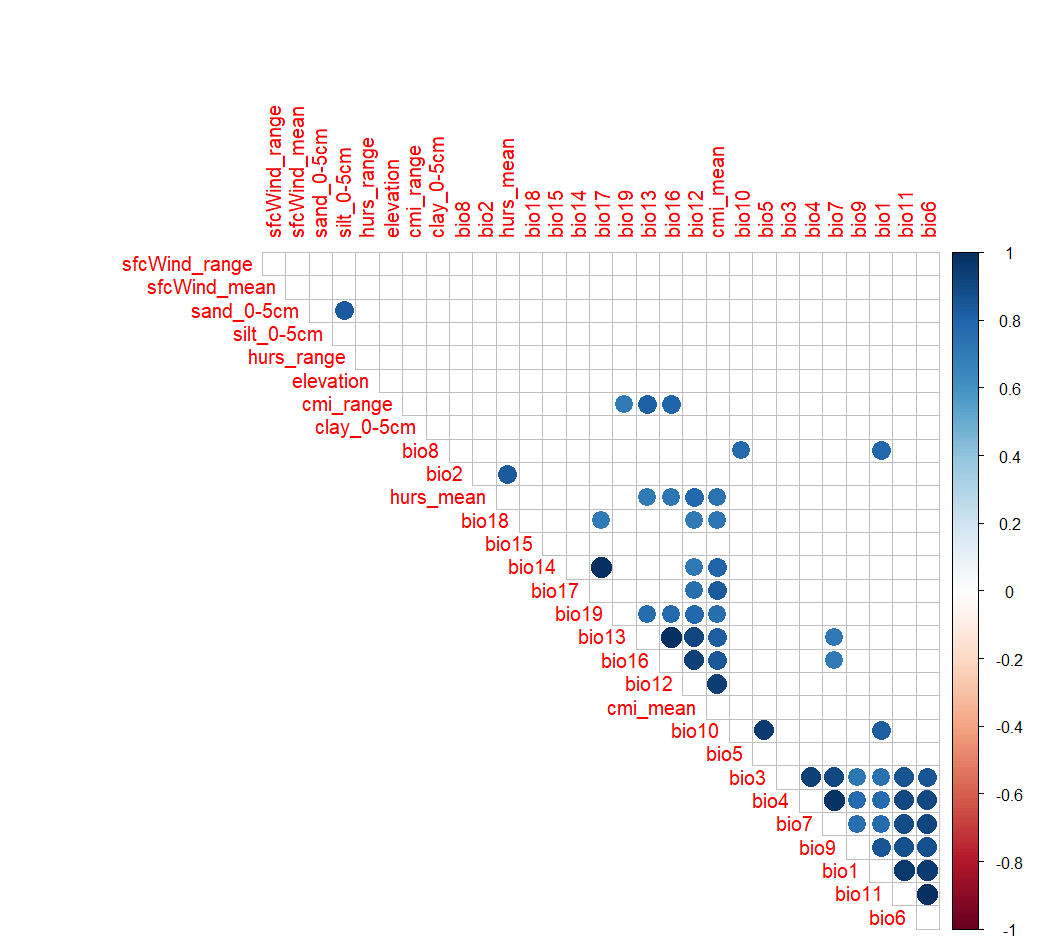
**hurs**: near-surface relative humidity (*only current conditions*)

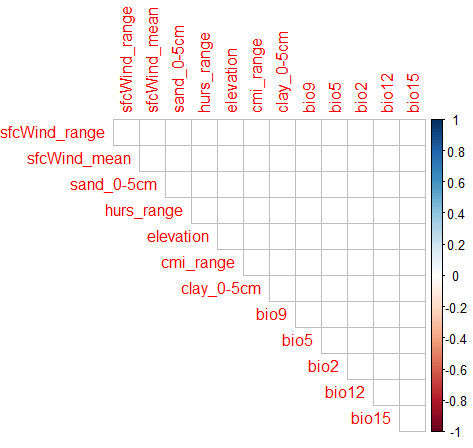
**sfcWind:** near-surface wind speed (only current conditions)

**Algorítimos**

Utilizaremos diferentes tipos de algoritmos e técnicas de modelagem devido à ausência de um único algoritmo que possa lidar com todas as condições de modelagem (Quiao et al. 2015; Norberg et al. 2019) .Serão utilizados os algoritmos, Support Vector Machine (**SVM**), Artificial Neural Network (**NNT**), Maximum Entropy (**MAXENT**), Gausian Process (**GP**), Random Forest (**RF**), Boosted Regression Trees (**BRT**), Modelos Lineares Generalizados (**GLM**), e Modelos Aditivos Generalizados (**GAM**).

Para os algoritmos que tem hiperparámetros como **o SVM, NNT, MAXENT, BRT, e RF** será utilizada a técnica “**busca por grid**” como método de seleção da combinação dos valores dos hiperparámetros que retorne a melhor performance.

****

****