

# GridDER: Grid Detection and Evaluation in R

Anônimo

**Abstract** Observações e coleções de organismos formam a base da nossa compreensão sobre a biodiversidade da Terra e são um recurso indispensável para inúmeros estudos. A informação geográfica é fundamental, servindo de elo entre os organismos e os ambientes em que residem. No entanto, a informação geográfica associada a esses registros é muitas vezes imprecisa, limitando a sua eficácia para pesquisas. Nessa linha, foram desenvolvidas várias soluções para identificar coordenadas errôneas e georreferenciados a centroides de países, estados ou municípios. Outra fonte de imprecisões proeminente, mas menos discutida e documentada, surge devido ao uso grades (em muitas regiões do mundo). Por isso, desenvolvemos o pacote R GridDER, para identificar registros de biodiversidade provenientes de sistemas de grades. Nossa ferramenta também estima o grau de heterogeneidade ambiental associado a essas grades, permitindo que os usuários tomem decisões informadas sobre como usar esses dados de ocorrência em estudos de mudanças globais. Mostramos que uma proporção significativa (~ 13,5%; 261 milhões) de registros no GBIF, a maior coleção de dados de coleta de história natural, são dados potencialmente em grade e demonstramos que nossa ferramenta pode identificar esses registros de forma confiável e quantificar as incertezas associadas. Portanto, o GridDER pode servir como uma ferramenta não apenas para triagem de dados georreferenciados em grade, mas para quantificar as incertezas geográficas e ambientais associadas a esses registros, que podem ser usados para informar modelos e análises que utilizam esses dados, incluindo aqueles referentes a mudanças globais.

**Palabras clave:** R package - biodiversity - occurrence data

## Métodos e fluxo de trabalho do pacote

Compilamos os seguintes metadados de sistemas de grade: nome do sistema de grade (se disponível), nome do país e código ISO-3166 (se um sistema de grade for aplicável a um ou vários países), o European Petroleum Survey Group (EPSG) do sistema de referência de coordenadas (CRS). O EPSG hospeda um banco de dados de CRSs, a resolução do sistema de grade em dimensões horizontais e verticais (tamanho e unidade) e a extensão espacial. Com os metadados compilados de sistemas de grade, simulamos sistemas de grade como vetores espaciais. Em resumo, desenvolvemos a função `grid_generation` para simular pontos regularmente espaçados representando os quatro vértices de grades num sistema de grades. Criamos também outras funções para simular sistemas de grade com base em ocorrências empíricas em grade. Isso se aplica a situações em que não há metadados disponíveis ou quando os metadados de um sistema de grade estão incompletos. Dentre os metadados utilizados para simular uma grade, o CRS é o atributo mais importante; quase sempre é verdade para um conjunto de dados espaciais. O CRS de um sistema de grade e ocorrências em grade são originalmente definidos pode se perder durante a agregação de dados, onde ocorrências de diferentes fontes com diferentes CRSs são comumente unificadas no mesmo CRS (por exemplo, WGS84).

Desenvolvemos a função `infer_crs` para inferir o CRS de um sistema de grade, usando coordenadas associadas a um sistema de grade desconhecido como entrada. Em situações em que sabe-se a unidade de distância espacial do CRS, mas resolução espacial é desconhecida, desenvolvemos uma função (`infer_resolution`) para inferir a resolução espacial com base nas ocorrências de entrada. Já a função `infer_origin` para calcular a origem das malhas quando o CRS e as resoluções espaciais são conhecidos. A origem é calculada como o mínimo  $x$  e  $y$  entre os pontos espaciais de entrada menos metade da resolução ao longo dos eixos  $x$  e  $y$ , respectivamente. A extensão espacial de um sistema de grade pode ser calculada pela função `infer_extent`. A extensão pode ser simplesmente a extensão espacial onde um CRS é definido, o limite do país ou a o delimitador das ocorrências em grade. Buffers adicionais, na unidade de resolução de grade, também podem ser adicionados junto com o sistema de grade, utilizando o parâmetro `flag_buffer` da função `grid_generation`. Também é possível calcular dois tipos de incerteza para cada grade do sistema: 1- incerteza ambiental com a função `assess_env_uncertainty` que calcula o desvio padrão de variáveis ambientais para de cada grade de um sistema de grades. O padrão do GridDER é o dado de elevação com 30 metros de resolução (NASDEM [https://developers.google.com/earth-engine/datasets/catalog/NASA\\_NASADEM\\_HGT\\_001](https://developers.google.com/earth-engine/datasets/catalog/NASA_NASADEM_HGT_001)), mas outras variáveis podem ser fornecidas pelo usuário; 2- incerteza espacial com função `assess_sp_uncertainty`, calcula a distância entre os centroides da grade e as coordenadas.

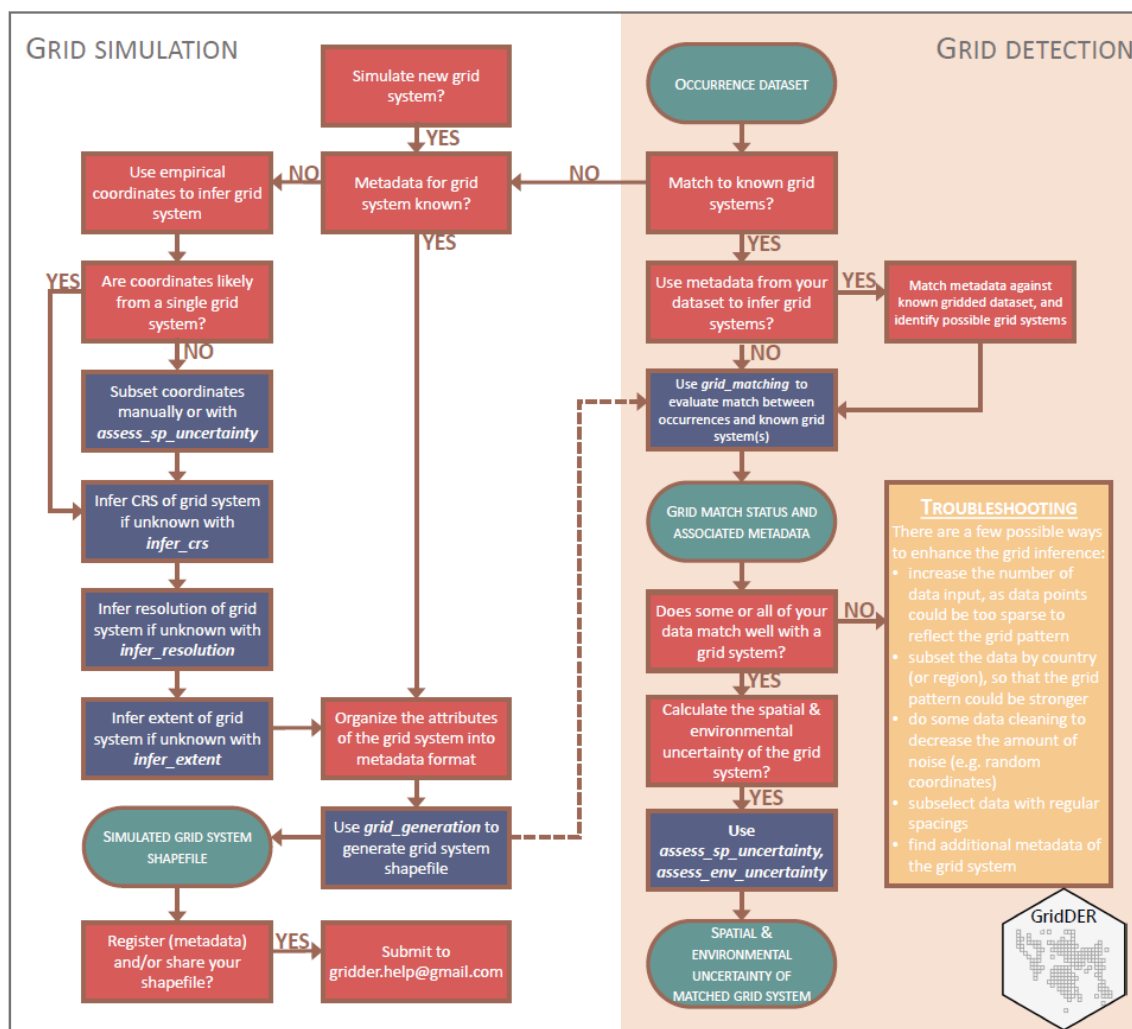


Figura 1: Fluxo de trabalho do pacote GridDER

O pacote ainda não está no CRAN, contudo, existe um preprint disponível e submetido ao periódico Global Ecology and Biogeography. Com as devidas revisões e contribuições, esperamos que em breve possamos submeter o pacote ao CRAN.