

Predicting climate-resilient habitats for *Vicia faba* L.: A data-driven approach to enhance multifunctional agroecosystems



Samaneh Samaei-Daryan¹, Martin Wiehle², Behnam Kamkar¹

¹Department of Agrotechnology, Faculty of Agriculture, Ferdowsi University of Mashhad, Iran
²Organic Plant Production and Agroecosystems Research in the Tropics and Subtropics, University of Kassel, Witzenhausen 37213, Germany

Introduction

- Legumes are strategic crops for sustainable, climate-resilient agriculture and plant protein supply.
- V. faba* is a protein-rich food/feed legume that also improves soil fertility through nitrogen fixation.

Research gap

- Climate change may shift the geographic suitability of *V. faba*, making it critical to identify climatically suitable areas for current and future crop planning.

Aim

- To assess current and future climatic suitability of *V. faba* using a robust, transferable species distribution modeling (SDM) approach.

SDM framework & Results

1 *V. faba* occurrence data processing

Raw occurrence records (n = 105,390)

↓

Data cleaning and filtering → Fig. 1.

↓

Spatial sampling bias assessment

Spatial autocorrelation

↓

KDE bias layer

Spatial thinning

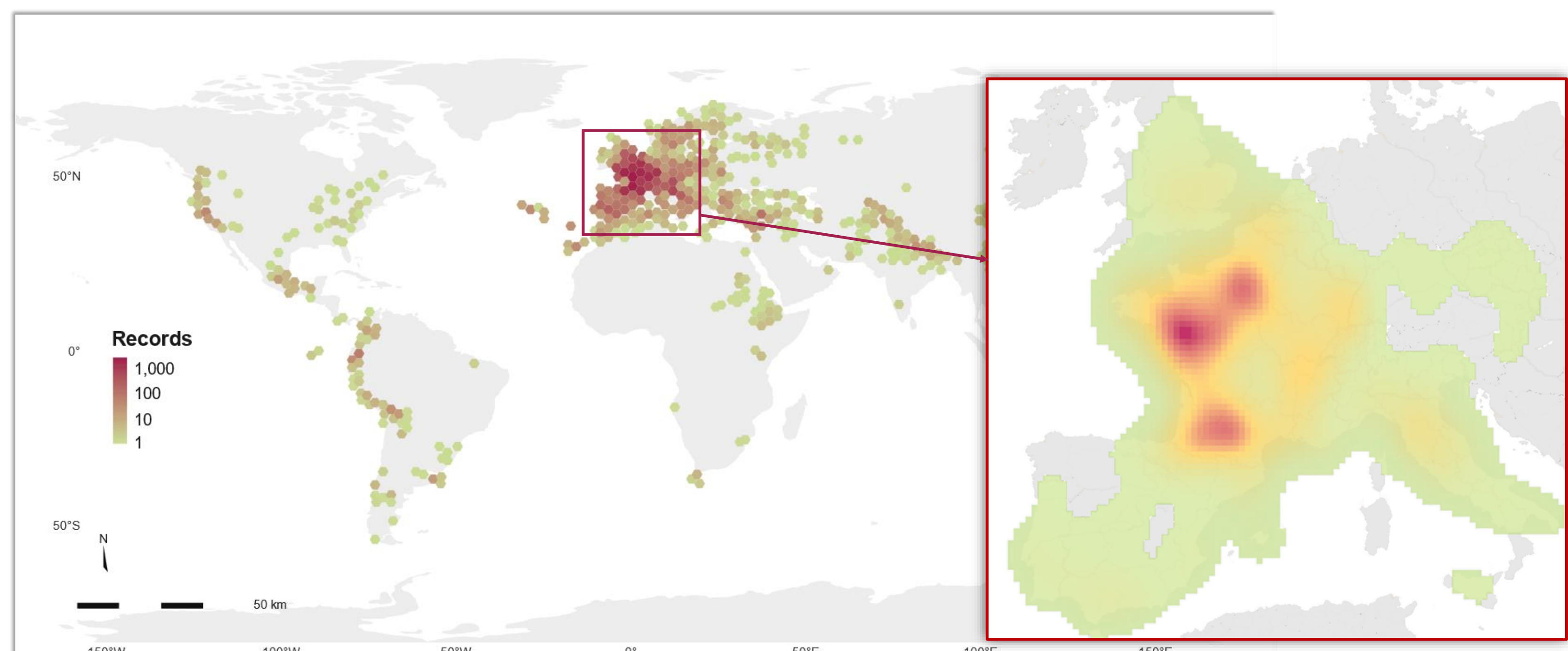


Fig. 1. Global distribution of cleaned *V. faba* occurrence records with a kernel density estimate of the main high-density cluster (Western Europe), revealing sampling bias and forming the basis for the bias-correction layer.

2 Environmental data processing

Literature-based identification of relevant variables for *V. faba*

↓

Candidate environmental variables (n = 26)

↓

Correlation analysis → Fig. 2

Multicollinearity → Table 1

Sources. WorldClim 2.1, SoilGrids v1 & v2, Wang et al. (2025).

High multicollinearity exists with a VIF value higher than 10.

Table 1. Variables with VIF > 10 were removed, leaving 17 variables retained for further analysis.

Category	Variables
Temperature	Bio2, Bio3, Bio4, Bio10
Topography	Elevation
Precipitation	Bio13, Bio14, Bio15
Soil	pH, Clay, Silt, OC, N, BD, CEC, Salinity, Depth

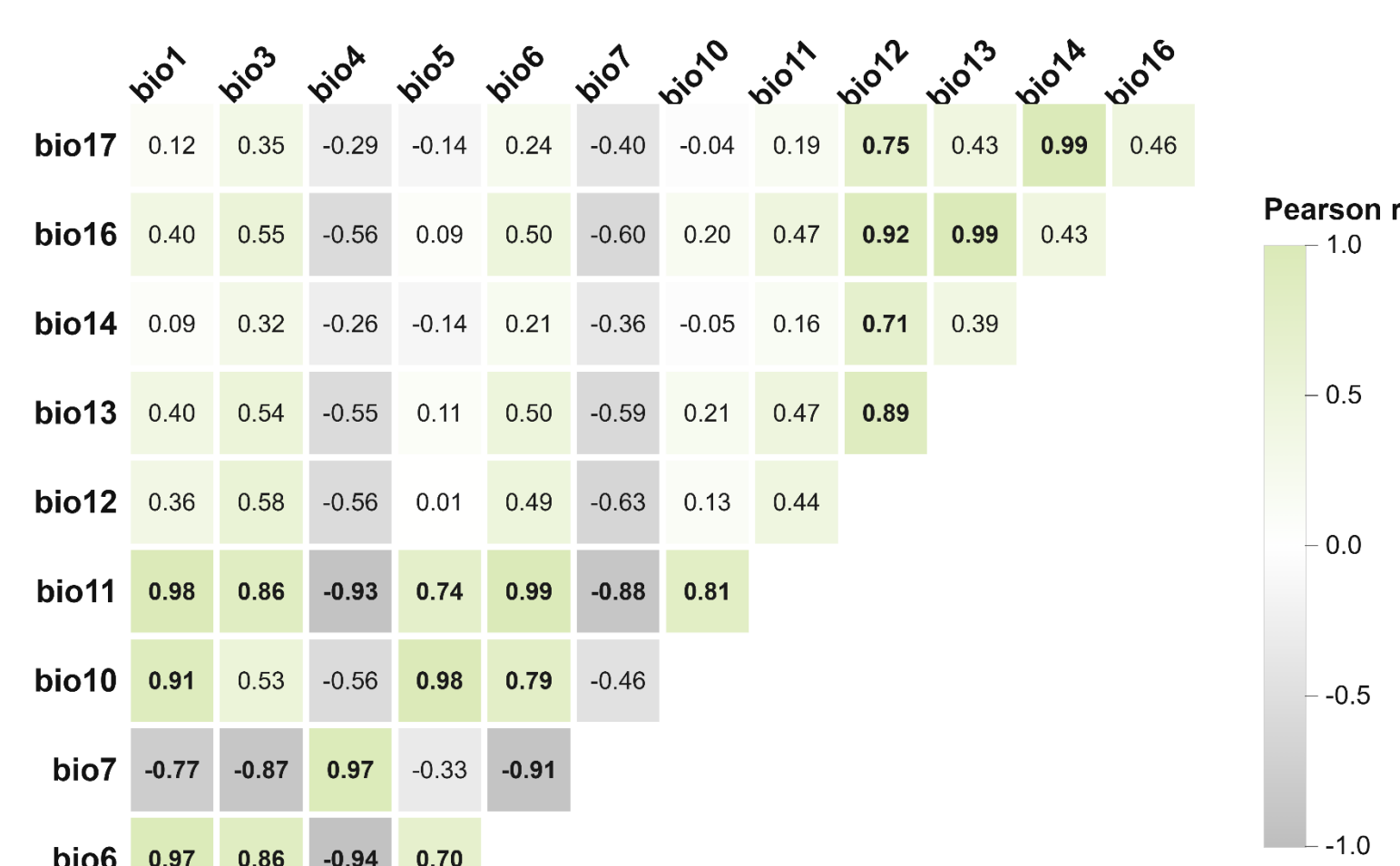


Fig. 2. Pearson correlation heatmap showing the strength and direction of correlations among the candidate variables.

3 Data partitioning for model evaluation

Cleaned occurrence

Selected variables

KDE bias layer

Background layer

↓

Spatial cross-validation, nearest-neighbour distance matching (kNNDM) → Fig. 3

❖ All analyses conducted in R.

Multiple clustering methods and fold counts (k = 2–5) were systematically compared before selecting k-means at k = 3.

This configuration minimized the nearest-neighbour distance mismatch (W) and showed the most stable performance across repeated seeds.

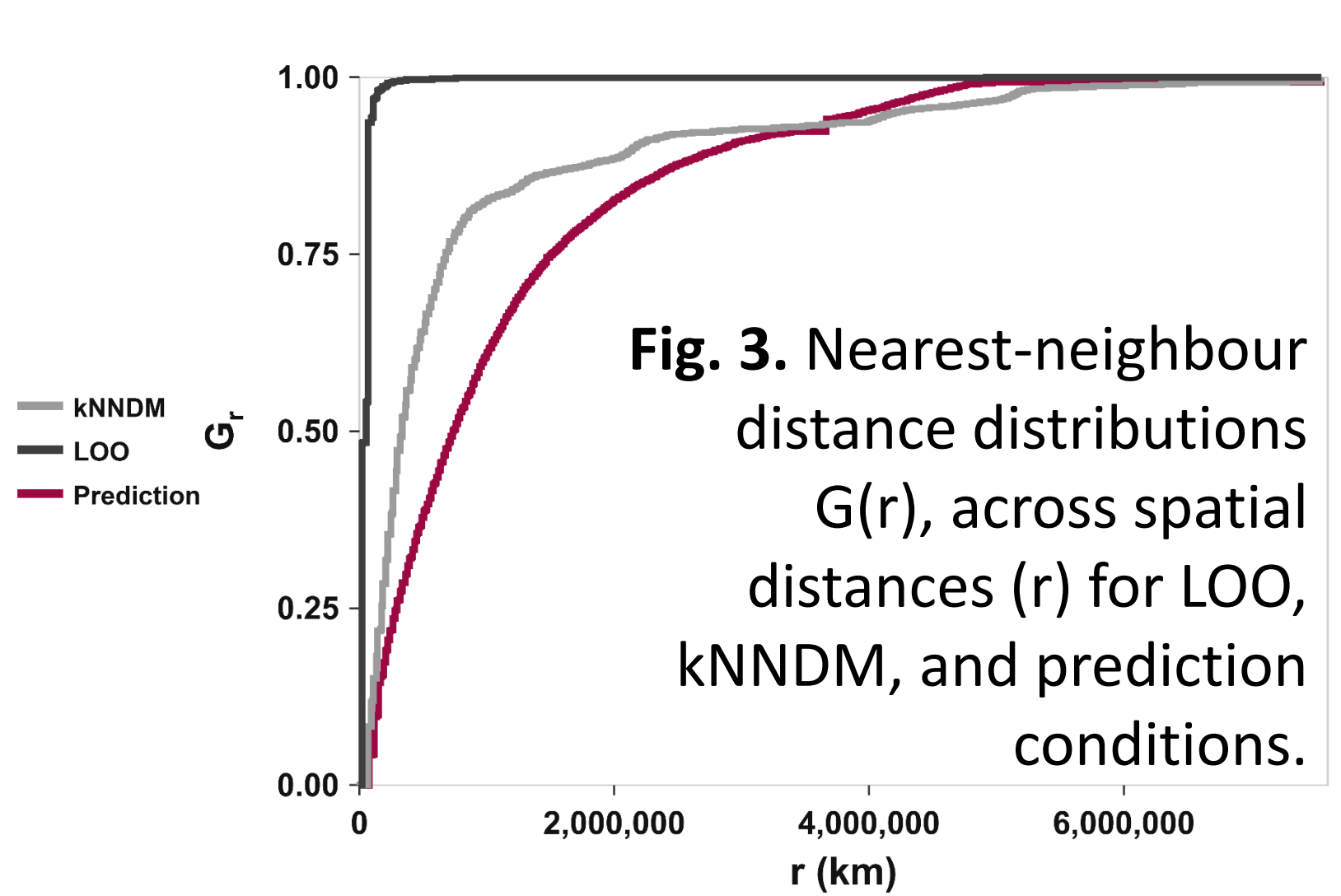


Fig. 3. Nearest-neighbour distance distributions G(r), across spatial distances (r) for LOO, kNNDM, and prediction conditions.

Conclusions

- The reproducible framework addresses key methodological challenges of applying SDMs to cultivated crops, offering a transferable foundation for land suitability assessment.
- Suitability maps and variable importance analyses can inform agricultural planning and *in situ* and *ex situ* conservation strategies for *V. faba*.