

FAST METHODS FOR FITTING LOG-GAUSSIAN COX PROCESS MODELS IN ECOLOGY

ELLIOT DOVERS 

(Received 28 November 2022; first published online 13 January 2023)

2020 Mathematics subject classification: primary 62P12; secondary 60G55, 92D40.

Keywords and phrases: automatic differentiation, ecology, fixed rank kriging, integrated nested Laplace approximation, log-Gaussian Cox Process, point patterns, spatial point process, species distribution model, variational approximation.

Log-Gaussian Cox processes (LGCPs) offer a framework for modelling point patterns that can accommodate latent effects. These latent effects can be used to account for missing predictors or other sources of clustering that could not be explained by a Poisson process. Such models are important in ecology where point patterns arise in the form of presence-only data—records of species' locations—and are used to construct Species Distribution Models (SDMs) as a function of environmental variables. Fitting LGCP models can be difficult and time consuming and, as a result, limits the ability of researchers to flexibly analyse presence-only data. We develop novel methodology and software for fitting LGCP models, as well as demonstrating how to incorporate presence-only and other data sources jointly into SDMs.

Fitting LGCPs quickly is challenging due to their intractable marginal likelihood which involves a high dimensional integral to account for the latent Gaussian field and leads to large spatial variance-covariance matrices. We address these challenges using a novel combination of variational approximation and reduced rank interpolation. Additionally, we implement automatic differentiation that enables us to obtain exact gradient information rapidly for computationally efficient optimisation and inference. We demonstrate the method's performance through both simulations and a real data application, with promising results in terms of computational speed and accuracy compared to that of existing approaches.

We then extend our novel method to combine presence-only data with that obtained through scientific surveys to improve SDMs in what is called data integration. We demonstrate scenarios in which sharing both the latent influence and the species'

Thesis submitted to the University of New South Wales in May 2021; degree approved on 13 October 2021; supervisors David Warton and Gordana Popovic.

© The Author(s), 2023. Published by Cambridge University Press on behalf of Australian Mathematical Publishing Association Inc.

response to the environment across each dataset can improve upon results achieved by modelling each individually, both via simulation and by using real data involving several species of flora in NSW, Australia.

We also illustrate the use of software developed to implement these advances via the freely available R package (<https://github.com/ElliottDovers/scampr>). The package allows users to fit likelihood-based LGCP to presence-only data swiftly and with a formula interface familiar to those in other regression-style modelling frameworks implemented on R.

ELLIOT DOVERS, School of Mathematics and Statistics,
University of New South Wales, Kensington, New South Wales 2052, Australia
e-mail: elliott.dovers@gmail.com