TEST (2018) 27:542–544
https://doi.org/10.1007/s11749-018-0606-2

DISCUSSION

Comments on: Some recent work on multivariate Gaussian Markov random fields

Miguel A. Martinez-Beneito[1,2]

Published online: 12 September 2018

Abstract

I provide comments on the article 'Some recent work on multivariate Gaussian Markov random fields' by Ying MacNab.

Keywords Multivariate disease mapping · Gaussian Markov random fields · Bayesian statistics · Coregionalization models

Comments

First of all I would like to start by congratulating the author for this impressive review paper. Much work has been published on the development of multivariate Gaussian Markov random fields (MGMRFs), following many different approaches, and all that work is not so easy to summarize. In my opinion the author has accomplished that valuable task very successfully.

One of the main contributions of the review paper is the proposed classification of the different approaches followed to build MGMRFs. The author has classified most of that previous work into conditional multivariate, conditional univariate and coregionalized models. This division seems very sensible and puts the previously published literature in order. Despite the interest of these three approaches, I have to admit my own particular preference for coregionalized models. In my opinion coregionalization models have several features that make them quite appealing and that make them a key element in multivariate spatial modelling. These advantages

This comment refers to the invited paper available at https://doi.org/10.1007/s11749-018-0605-3.

✉ Miguel A. Martinez-Beneito
martinez_mig@gva.es

[1] Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana (FISABIO), Valencia, Spain

[2] CIBER de Epidemiología y Salud Pública, Madrid, Spain

could be summarized as computational convenience and validity by construction. I elaborate further on these issues in the following.

Coregionalization models have been shown to be outstanding tools in computational terms for several reasons. First, they are built as simple linear combinations of underlying spatial patterns, which is a very sensible and computationally economical idea. Second, most of the operations used for structuring dependence under this approach comprise Kronecker products or block matrix operations, which can be easily expressed as simple matrix products (Martinez-Beneito 2013). This has allowed the coregionalization framework to be successfully coded and used even in `WinBUGS`, which contains a very limited number of mathematical tools, in particular for matrix operations. Moreover, coregionalization models, combined with orthogonal transformations to avoid the effect of the order in which the diseases are considered, have been shown to yield significant advantages in computational terms. Specifically, this combination avoids the need to perform Cholesky decompositions of the covariance matrix between diseases or to deal with orthogonal matrices that can be computationally expensive or problematic for MCMC convergence. As mentioned in MacNab's paper that combination has made it possible to perform multivariate disease mapping studies of even 21 diseases in lattices of 540 spatial units (Botella-Rocamora et al. 2015). Moreover, these computational advantages have also allowed coregionalization models to be applied for the joint study of several diseases, sexes, time periods, etc., altogether (Martinez-Beneito et al. 2017). To our knowledge, this kind of study, or studies with so many dimensions, has not yet been conducted using conditional univariate or multivariate approaches. But beyond these computational benefits, coregionalization casts MGMRFs within a matrix algebraic framework, which makes it possible to introduce all matrix theory and its computational advantages into the practice of this field.

As mentioned, the second main advantage of coregionalization models is that they are valid by construction. This may seem a minor advantage since sufficient conditions are usually set for conditionally specified models, but this is not so small. MGMRFs models are usually constrained according to some restriction, usually diagonal dominance of the precision matrix, which guarantees the fulfilment of such sufficient conditions. Nevertheless, no idea is usually had of the number of models excluded by restricting ourselves to models fulfilling that sufficient condition. As a consequence, an important part of the corresponding family of valid MGMRFs could be missed by setting a tight sufficient condition on their validity. Moreover, the extreme general scope of conditionally specified models makes them capable of reproducing practically any model that we could think of, even lots of senseless or plainly invalid models. Thus, the real challenge of conditionally specified models is to define models that are, at the same time, valid, computationally feasible and sensible models. Coregionalization models yield proposals with all these three features, which is a non-negligible advantage.

In any case, I admit that, as pointed out in MacNab's paper, coregionalization models seem to be somewhat more restrictive than multivariate conditionally defined models. Namely, multivariate models with asymmetric cross-spatial dependence may be defined for both conditional univariate (Sain et al. 2011) and multivariate (Greco and Trivisano 2009) models. However, asymmetric models have not yet been proposed

within the coregionalization framework, which seems to be a limitation compared to conditional models. Nevertheless, as pointed out in MacNab's paper, this asymmetry also implies an asymmetric behaviour between spatial units or, in other words, induces label dependence for them. Thus, region i may have a different effect on region k depending on whether $i < k$ or $i > k$, when the labelling of spatial units is typically arbitrary. Therefore, an obvious question arises here: would we want to have a degree of freedom in our studies that basically introduces an uncontrolled arbitrary effect on them? Basically, for asymmetric models the results could be somewhat different for two analysts using different orderings for either their diseases or their spatial units. I am not so sure, and this is my particular opinion, we want to have that degree of arbitrariness in our analyses.

Finally, MacNab's paper leaves, as I see it, some additional open questions that are sure to guide new research, which makes her contribution particularly valuable. For example, can any of the models of any of MacNab's three modelling sets be reproduced within the rest of the approaches? Or, in other words, are those three different sets of models mostly equivalent or are they instead complementary in the sense that the models reproduced in any of them cannot be generally reproduced by the other approaches? Moreover, if these sets of models were really distinct, what particular features do the models of one of these approaches that cannot be reproduced by the rest show? That is to say, which specific modelling features can be reproduced for any of those approaches that cannot be reproduced for the rest? All these question are really thought-provoking and their answers may yield important insights on the different approaches of an apparently fragmented field, which would be enriched by a general overview that put all three approaches in common.

References

Botella-Rocamora P, Martinez-Beneito MA, Banerjee S (2015) A unifying modeling framework for highly multivariate disease mapping. Stat Med 34(9):1548–1559

Greco FP, Trivisano C (2009) A multivariate CAR model for improving the estimation of relative risks. Stat Med 28(12):1707–1724

Martinez-Beneito MA (2013) A general modelling framework for multivariate disease mapping. Biometrika 100(3):539–553

Martinez-Beneito MA, Botella-Rocamora P, Banerjee S (2017) Towards a multidimensional approach to bayesian disease mapping. Bayesian Anal 12:239–259

Sain S, Furrer R, Cressie N (2011) A spatial analysis of multivariate output from regional climate models. Ann Appl Stat 5(1):150–175