

Document downloaded from:

[\[http://redivia.gva.es/handle/20.500.11939/6180\]](http://redivia.gva.es/handle/20.500.11939/6180)

This paper must be cited as:

[Martínez-Minaya, J., Conesa, D., López-Quílez, A., & Vicent, A. (2018). Spatial and climatic factors associated with the geographical distribution of citrus black spot disease in South Africa. A Bayesian latent Gaussian model approach. *European journal of plant pathology* 151: 991-1007.]

**ivia**  
Institut Valencià  
d'Investigacions Agràries

The final publication is available at

[\[http://dx.doi.org/10.1007/s10658-018-1435-6\]](http://dx.doi.org/10.1007/s10658-018-1435-6)

Copyright [Springer]

1 **Spatial and climatic factors associated with the geographical distribution of citrus black spot disease**  
2 **in South Africa. A Bayesian latent Gaussian model approach**

3

4 **Joaquín Martínez-Minaya<sup>2</sup>, David Conesa<sup>2</sup>, Antonio López-Quílez<sup>2</sup>, Antonio Vicent<sup>1</sup>**

5

6<sup>1</sup> Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA),  
7 Moncada, 46113 Valencia, Spain.

8

9<sup>2</sup> Departament d'Estadística i Investigació Operativa, Universitat de València, C/ Dr. Moliner 50, Burjassot,  
10 46100 Valencia, Spain.

11

12 Corresponding author:

13 A. Vicent

14 E-mail [vicent\\_antciv@gva.es](mailto:vicent_antciv@gva.es)

15 Tel. (+34) 963424078

16 Fax. (+34) 963424001

17

18 **Abstract**

19

20 Citrus black spot (CBS), caused by *Phyllosticta citricarpa*, is the main fungal disease affecting this crop  
21 and quarantine measures have been implemented. The role of climate as a limiting factor for the  
22 establishment and spread of CBS to new areas has been debated, but previous studies did not address the  
23 effects of spatial factors in the geographic distribution of the disease. The effects of climatic and spatial  
24 factors were studied using South Africa as a case study, due to its diversity of climates within citrus-  
25 growing regions. Georeferenced datasets of CBS presence/absence in citrus areas were assembled for two  
26 stages of the epidemic: 1950 and 2014. Climatic variables were obtained from the WorldClim database.  
27 Moran's I and Geary's C analyses indicated that CBS distribution data presented significant spatial  
28 autocorrelation, particularly in 2014. Collinearity was detected among climatic variables. Spatial logistic  
29 regressions, particular case of latent Gaussian models, were fitted to CBS presence/absence in 1950 or  
30 2014 with the Integrated Nested Laplace Approximation methodology. Principal components (PCs) or pre-  
31 selection of climatic variables based on their correlation coefficients were used to cope with collinearity.

32Spatial effects were incorporated with a geostatistical term. In general, the models indicated a positive  
33relationship between CBS presence and climatic variables or PCs associated with warm temperatures and  
34high precipitation. Nevertheless, in 1950, models that also included a spatial effect outperformed those  
35with climatic variables only. Problems of model convergence were detected in 2014 due to the strong  
36spatial structure of CBS distribution data. The consequences of ignoring spatial dependence to estimate the  
37potential geographical range of CBS are discussed.

38

39**Keywords** *Guignardia citricarpa*, geostatistics, INLA, biogeography, risk assessment

40

41

42Citrus black spot (CBS) disease, caused by the fungus *Phyllosticta citricarpa* (McAlpine) van der Aa  
43(synonym *Guignardia citricarpa* Kiely), is the most important fungal disease affecting this crop  
44worldwide. The pathogen causes external blemishes on the fruit rind and induces premature fruit drop,  
45resulting in serious quality and yield losses (Martínez-Minaya et al. 2015). When complementary mating  
46types are present, the pathogen reproduces through sexual spores (ascospores) formed in the leaf litter after  
47a maturation process driven mainly by temperature and moisture (McOnie 1964b; Tran et al. 2017). Once  
48mature, ascospores are discharged from the leaf litter and disseminated by air currents (McOnie 1964b).  
49Ascospores infect susceptible fruit, twigs and leaves in the presence of moisture and conducive  
50temperatures. The pathogen also reproduces by asexual spores (conidia), which are rain-splashed mainly  
51from lesions in citrus fruit and twigs (Perryman et al. 2014). Ascospores have been deemed as the main  
52source of inoculum in South Africa (McOnie 1964b), but studies in Brazil and Florida (USA) have  
53suggested that conidia are epidemiologically important under certain conditions (Hendricks et al. 2017;  
54Spósito et al. 2008; Wang et al. 2016). The application of chemical fungicides is generally needed for CBS  
55control (Makowski et al. 2014), resulting in increased environmental and economic costs of citrus  
56production.

57 The CBS disease is currently present in important citrus-growing regions of Australia, Asia, Africa  
58and America. Quarantine measures have been established by several countries, such as South Africa, USA,  
59Japan and the European Union (EU), to prevent the entry of *P. citricarpa* into areas that are still free of the  
60pathogen (Martínez-Minaya et al. 2015). According to the International Plant Protection Convention  
61(IPPC) and the World Trade Organization (WTO), phytosanitary regulations should be based on a

62scientific pest risk analysis (PRA). PRAs are based on standardized protocols aimed at estimating the  
63likelihood of disease introduction (i.e. entry and establishment) and subsequent spread in order to devise  
64the most efficient options as regards risk reduction. Maps describing host availability and climatic  
65suitability for disease development are a key component of PRAs to set bounds on potential introductions  
66into new areas (Venette et al. 2010).

67 Paul et al. (2005) estimated the potential global geographical range of CBS using CLIMEX. They  
68concluded that climates in the Mediterranean Basin were not suitable for CBS and, therefore, phytosanitary  
69measures for *P. citricarpa* in the EU were not necessary. However, Paul et al. (2005) were not able to  
70predict the presence of the disease in the arid citrus-growing areas of the Eastern Cape province in South  
71Africa, where CBS is endemic. In a subsequent study, Yonow et al. (2013) modified the parameters of Paul  
72et al. (2005) allowing CLIMEX to predict the presence of CBS in this region. Using a new set of CLIMEX  
73parameters, Er et al. (2013) predicted climatic suitability for CBS in arid areas of Mediterranean-type  
74climates in California (USA).

75 A mechanistic (process-based) generic infection model (Magarey et al. 2005) was used to obtain maps  
76of climate suitability for CBS. This generic infection model consisted of parameters for temperature and  
77wetness duration, and it was specifically developed for exotic pathogens, like *P. citricarpa*, on which there  
78is little biological information. One study using this model concluded that the climates of the EU cannot be  
79considered as unsuitable for the establishment of *P. citricarpa* (EFSA 2008), whereas another study  
80indicated that CBS was not expected to have an impact in areas with commercial citrus production in  
81Europe (Magarey et al. 2011). Nevertheless, due to the paucity of biological information available for *P.*  
82*citricarpa*, the results obtained with process-based models were highly uncertain (EFSA 2014).

83 Empirical models for *Phyllosticta* spp. ascospore maturation and release (Fourie et al. 2013) have been  
84combined with the generic infection model so as to restrict predictions only to the periods of potential  
85ascospore availability. One study using these models indicated that climatic conditions in many EU citrus-  
86growing areas were suitable for CBS (EFSA 2014), whereas another suggested that only a few isolated  
87locations in Europe have a low to marginal risk of *P. citricarpa* establishment (Magarey et al. 2015).  
88However, in this latter study, infection events were dramatically diminished, only those associated with

89rains being considered. Nevertheless, as indicated above, large uncertainties have been associated with  
90these models and their applications (EFSA 2016).

91 Correlative statistical models are widely used in different areas of biogeography, such as conservation,  
92resource management, global warming and biological invasions (Franklin 2009). However, their use for  
93risk assessment in plant pathology is still limited and few studies are available for diseases caused by fungi  
94or oomycetes (Elith et al. 2013; Meentemeyer et al. 2008; Narouei-Khandan et al. 2017). Typically,  
95correlative species distribution models explore the relationships between species occurrences and climatic  
96variables to produce maps of predicted distributions of the target organisms. Without enough biological  
97and epidemiological information for process-based models, correlative species distribution models may  
98help to identify climatic variables that are associated with CBS and, therefore, demarcate locations that  
99would allow disease establishment.

100 Several statistical methods are used for species distribution modelling based on presence/absence data,  
101such as generalized linear models (GLM) and generalized additive models (GAM) (Franklin 2009). In  
102many cases, species distribution models are used without considering the spatial dependence of the data,  
103assuming that the geographical range is only driven by climate and the disease is in equilibrium with these  
104factors. However, this assumption is often violated when disease spread is constrained due to dispersal  
105barriers and/or absence of susceptible host plants. Moreover, ignoring spatial autocorrelation may lead to  
106inaccurate parameter estimates, inadequate quantification of uncertainty, and thus poor predictive power.  
107With spatially explicit hierarchical Bayesian models it is possible to introduce the effect of spatial  
108dependence (Latimer et al. 2006). These complex models have usually been fitted with Markov chain  
109Monte Carlo (MCMC) methods that are computationally costly, especially for large spatial datasets. In the  
110specific case of latent Gaussian models, Approximate Bayesian inference with integrated nested Laplace  
111approximations (INLA) is a much faster and computationally efficient alternative to MCMC (Lindgren et  
112al. 2011; Rue et al. 2009).

113 The main objective of this study was to analyse the spatial and climate effects that influence the  
114probability of CBS presence in South Africa. South Africa was used here as a case study because it is the  
115only country with commercial citrus production under ten climate types, covering a wide range of  
116environmental conditions. This information will help risk managers to better understand the factors  
117associated with the potential establishment and spread of CBS into new areas.

118

## 119 **Materials and methods**

120

121 Datasets

122

123 Spatially gridded datasets including presence and absence of CBS in citrus-growing areas in South  
124 Africa were assembled for 1950 and 2014. A raster layer (5' arc min resolution) of citrus distribution in  
125 South Africa was generated by a mapping specialist from the map of citrus trees in South Africa published  
126 by Powell (1930), based on the census carried out in 1927 and restricted to the boundaries of South Africa  
127 (Fig. 1a). Grid cells with citrus were classified as CBS-present ( $n = 28$ ) and CBS-absent ( $n = 776$ ) based on  
128 the survey included in Appendix 2 of Wager (1952), which was conducted from 1940 to 1950. This latter  
129 year was used to denote the dataset. Since the coexistence of pathogenic and non-pathogenic species of  
130 *Phyllosticta* in citrus was not discovered until a decade later (McOnie 1964a), reports of the pathogen in  
131 absence of CBS symptoms were excluded from Wager (1952).

132 For 2014, the spatially gridded South African national land-cover (NLC) dataset 2013-2014 was used  
133 (DEA 2015), but including only those grid cells of the class “cultivated commercial permanent orchards”  
134 enclosed within the citrus areas of the map by Paul (2005) and its subsequent updates (Anonymous 2014;  
135 Yonow et al. 2013). A raster layer of 5' arc min resolution was assembled with CBS-present ( $n = 620$ ) and  
136 CBS-absent ( $n = 313$ ) grid cells (Fig. 1b) (Anonymous 2014; DEA 2015; Paul 2005; Yonow et al. 2013).  
137 Phytosanitary barriers for the internal movement of citrus plants in South Africa (Fig. 1b) were gathered  
138 from official governmental regulations (Anonymous 1984, 2002), as reviewed by Martínez-Minaya et al.  
139 (2015).

140 Spatially gridded climatic data (5' arc min resolution) from South Africa were acquired from the  
141 WorldClim database, which included mean monthly values for the period 1950-2000 (Hijmans et al. 2005).  
142 In addition to the 19 climatic variables available in WorldClim, precipitation from October to January and  
143 accumulated degrees (ADD) from July to October with (i.e. average of  $T_{max}$  and  $T_{min}$  for each month with  
144  $T_{base} = 10^{\circ}\text{C}$ ) were also calculated (Table 1). The coordinate system WGS84 was used in all spatially  
145 gridded datasets with the raster package for R 3.2.5 (Hijmans 2014; R Core Team 2016).

146

147 Spatial autocorrelation, collinearity and PCA

148

149 To test the hypothesis that the response variable (i.e. CBS presence/absence) occurs at random among  
150 citrus grid cells, Moran's *I* and Geary's *C* analyses of spatial autocorrelation were used (Plant 2012). For  
151 each dataset, 1950 and 2014, CBS-present citrus grid cells were coded with a 1 and CBS-absent citrus grid  
152 cells were coded with a 0. Grid cells without citrus were not considered in the analyses. Both indices were  
153 calculated by contiguity and at increasing distances from 20 to 900 km. Moran's *I* values range from -1,  
154 indicating perfect dispersion, to 1 indicating perfect correlation (i.e. clustering). The expected value of  
155 Moran's *I* in the absence of significant spatial autocorrelation is around 0. The value of Geary's *C* is 1 in  
156 the absence of spatial autocorrelation and approaches 0 for strong autocorrelation.

157 Pearson's correlation coefficient was used to detect collinearity among the 21 climatic explanatory  
158 variables included in the analysis. Pairwise correlations were classified as  $|r| \leq 0.7$  or  $|r| > 0.7$  according to  
159 Dormann et al. (2012a), who proposed this threshold of correlation as an appropriate indicator for when  
160 collinearity begins to severely distort model estimation and subsequent prediction.

161 Principal component analysis (PCA) was used to obtain independent linear combinations of 20  
162 climatic variables in order to summarize most of the variation in each dataset. The climatic variable  
163 temperature annual range was not taken into account in the PCA, because it is a linear combination of the  
164 variables maximum temperature of the warmest month and minimum temperature of the coldest month  
165 (Table 1). Principal components (PCs) were extracted sequentially according to the amount of variation  
166 explained. The relationship between the individual variables and the extracted PCs was expressed by a  
167 Varimax rotated component matrix with Kaiser's normalization. Values approaching unity indicated a  
168 greater contribution of the variable to the component (Chatfield and Collins 2013). In addition, two 95%  
169 confidence ellipses were plotted for pairwise PCs, one for CBS-present grid cells and another for CBS-  
170 absent grid cells (Johnson and Wichern 2002).

171

172 Models

173

174 A Bayesian hierarchical spatial approach was used to model the variation in the proportion of the  
175 presence. This approach can be considered as a spatial extension of a generalized linear model in the sense  
176 that a stochastic spatial effect is added to the linear predictor. Note also that this approach is highly suitable

177for situations in which data are observed at continuous locations occurring within a defined spatial domain.  
178Nevertheless, the main interest when dealing with this kind of model is to predict the response in  
179unsampled locations, usually known as kriging in honour of Krige's (Krige 1951) seminal work. From the  
180Bayesian point of view, this prediction can be performed via predictive distributions that easily allow the  
181incorporation of uncertainty within the model parameters.

182 In line with Muñoz et al. (2013), let  $Z_i$  be the binary response variable representing the presence (1) or  
183absence (0) of CBS at location  $i$ . Then, its conditional distribution is  $Z_i \mid \pi_i \sim \text{Ber}(\pi_i)$ ,  $\pi_i$  being the  
184probability of CBS presence at location  $i$ . As usual with GLMs, the mean of the response variable was  
185linked to the linear predictor and to the stochastic spatial effect by means of the logit link function defined

186as  $\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ . In particular,

$$187 \quad \text{logit}(\pi_i) = X_i \beta + W_i,$$

188 $\beta$  being the regression coefficients vector,  $X$  the covariates matrix and  $W_i$  the spatially structured random  
189effect. The geostatistical term  $W$  was assumed to be a multivariate Gaussian distribution whose covariance  
190matrix  $\sigma_w^2 H(\varphi)$  depends on the distance between locations, and with hyperparameters  $\sigma_w^2$  and  $\varphi$   
191representing the variance and range of the geostatistical term, respectively. Once the model had been  
192determined, posterior distributions had to be obtained. As in the Bayesian framework, parameters were  
193treated as random variables and prior knowledge had to be incorporated using the corresponding prior  
194distributions. These priors were specified jointly with random effects, the final level of the Bayesian  
195hierarchical model being the expression of the prior knowledge about the hyperparameters.

196 When dealing with Bayesian hierarchical models, posterior distributions for the parameters and  
197hyperparameters do not usually have any analytic expression, therefore numerical approximations are  
198needed. In the particular case of latent Gaussian models, INLA is a computationally efficient alternative to  
199MCMC. Latent Gaussian models are a particular case of the Structured Additive Regression (STAR)  
200models, where the mean of the response variable is linked to a structured predictor that accounts for the  
201effects of various covariates in an additive way. The prior knowledge of the additive predictor is expressed  
202using Gaussian prior distributions. In this context, all the latent Gaussian variables can be seen as  
203components of a vector known as the latent Gaussian Field (Rue et al. 2009).

204 To fit and predict the particular case of continuously indexed Gaussian Fields with INLA, as in our  
 205 case,  $W$ , an additional module is required. Lindgren et al. (2011) proposed an explicit link between  
 206 Gaussian Markov Random Fields (Rue and Held 2005) and continuous Gaussian Fields with a Matern  
 207 covariance structure via a weak solution to a stochastic partial differential equation (SPDE). Under this  
 208 approximation, the geostatistical spatial term is reparameterized as follows,

$$209 W \sim N(0, Q(\kappa, \tau)),$$

210 depending on two different parameters,  $\kappa$  and  $\tau$ , determining the range of the effect and the total variance,

211 respectively. More precisely, the range is approximately  $\varphi = \sqrt{\frac{8}{\kappa}}$  and the variance is  $\sigma_w^2 = \frac{1}{4\pi\kappa^2\tau^2}$

212 (Lindgren et al. 2011).

213 As mentioned above, the final step is to specify the prior distributions for the parameters and  
 214 hyperparameters. Normal vague priors with mean 0 and precision  $10^{-4}$  were used for the regression  
 215 coefficients vector. Although internally INLA works with  $\kappa$  and  $\tau$ , priors for the geostatistical term were  
 216 specified in terms of  $\varphi$  and  $\sigma_w$  using the reparameterizations  $\log(\varphi)$  and  $\log(\sigma_w)$  as independent  
 217 Gaussian distributions (Lindgren and Rue 2015).

218 To conclude, the full model was stated as follows:

$$\begin{aligned} Z_i \vee \pi_i &\sim \text{Ber}(\pi_i) \\ \text{logit}(\pi_i) &= X_i\beta + W_i, \\ 219 \beta_j &\sim N(0, 10^{-4}), W \sim N(0, Q(\varphi, \sigma_w)), \\ \log(\varphi) &\sim N(m_\varphi, q_\varphi), \\ \log(\sigma_w) &\sim N(m_{\sigma_w}, q_{\sigma_w}), \end{aligned}$$

220 where  $m_\varphi$  was chosen automatically such that the range of the field was about 20% of the diameter of the  
 221 region, and  $m_{\sigma_w}$  was chosen so that the corresponding variance of the field was 1 (in particular,  
 222  $m_\varphi = 1.476$  and  $m_{\sigma_w} = 0$ ). Finally, the precisions of the prior distributions for  $\log(\varphi)$  and  $\log(\sigma_w)$  were  
 223  $q_\varphi = 0.1$  and  $q_{\sigma_w} = 0.1$ .

224 Models including a selection of climatic explanatory variables with  $|r| \leq 0.7$  or PCs were fitted to the  
 225 response variable (CBS presence/absence). Models covering all possible combinations of climatic  
 226 explanatory variables with  $|r| \leq 0.7$  or PCs were compared using the Watanabe Akaike Information

227Criterion (WAIC), which uses the posterior densities more effectively than the traditional Deviance  
228Information Criterion (Gelman et al. 2014; Watanabe 2010). The models including climatic explanatory  
229variables or PCs displaying the lowest WAIC were selected. The geostatistical spatial term was  
230incorporated into these models as described above and the corresponding WAIC was calculated.

231 A validation dataset with CBS-present ( $n = 385$ ) and CBS-absent ( $n = 259$ ) grid cells (Fig. S1) was  
232assembled by random sampling without replacement from the 2014 dataset, but excluding those grid cells  
233used for model development in 1950. Receiver operating characteristic (ROC) curve analysis was used to  
234evaluate the predictive ability of the models selected for the 1950 dataset. CBS presence/absence was  
235considered as the binary classification variable. The mean of the predictive posterior distribution of  $\pi_i$   
236obtained with each model was evaluated as a continuous estimator of this binary classification variable.  
237ROC curves showed the proportion of correctly classified absences (specificity) in the x-axis and the  
238proportion of correctly classified presences (sensitivity) in the y-axis as the continuous variable moved  
239over its range of values (i.e. from 0 to 1). The area under del ROC curve (AUC) was calculated by  
240trapezoids using the pROC package for R (Robin et al. 2011).

241

## 242Results

243

244Spatial autocorrelation, collinearity and PCA

245

246 Moran's  $I$  and Geary's  $C$  analyses indicated the presence of significant spatial autocorrelation ( $P <$   
2470.0001) in CBS distribution data in 1950 and 2014 (Fig. 2). Both indices showed that spatial  
248autocorrelation was stronger in 2014 than in 1950. In 1950, Moran's  $I$  was highest from contiguity to 50  
249km, with a maximum of 0.33. Spatial autocorrelation decreased with distance and values of Moran's  $I$  close  
250to zero, approaching a random spatial pattern, were obtained from 600 km onwards. In 2014, values of  
251Moran's  $I$  equal to one (indicating perfect correlation) were obtained from contiguity and distances  
252between 20 and 180 km, with values higher than 0.79 from 190 to 900 km. In 1950, the lowest value of  
253Geary's  $C$  was 0.69 for contiguity and values close to one, indicating an absence of spatial autocorrelation,  
254were obtained with distances greater than 600 km. In 2014, values of Geary's  $C$  were lower than 0.21 in all  
255cases.

256 A high degree of collinearity was detected among the climatic variables, with 189 out of a total of 210  
257 pairwise correlations being significant ( $P < 0.05$ ) in 1950 (Fig. S2). Pairwise correlations with  $|r| > 0.7$  were  
258 detected: 17 among the temperature variables, 13 among the precipitation variables and only 3 among the  
259 temperature and precipitation variables. In 2014, 193 pairwise correlations were significant ( $P < 0.05$ ).  
260 Those with  $|r| > 0.7$  were 17 among the temperature variables, 13 among the precipitation variables and  
261 only 3 among the temperature and precipitation variables (Fig. S3).

262 Three PCs were extracted from the 1950 dataset, explaining 85.4% of the variability, with PC1 =  
263 37.4%, PC2 = 25.9% and PC3 = 22.1%. The temperature variables with the most influence in PC1 were the  
264 mean diurnal range and temperature seasonality with positive coefficients of 0.932 and 0.887, respectively  
265 (Table 1). The precipitation of the driest month and the driest quarter made a negative contribution to PC1,  
266 with coefficients of -0.876 and -0.879, respectively. Precipitation seasonality made a positive contribution  
267 to PC1, with a coefficient of 0.719. When plotted onto the map of South Africa, the lowest values of PC1  
268 coincided mainly with the Indian Ocean coastal areas (Fig. 3a). Temperature variables did not contribute  
269 much to PC2 (Table 1). Annual precipitation, precipitation of the wettest month and quarter made a strong  
270 positive contribution to PC2, with coefficients greater than 0.89. Precipitation in the warmest quarter and  
271 from October to January also made a strong positive contribution to PC2, with coefficients greater than  
272 0.97. The highest values of PC2 were obtained in the eastern half of South Africa (Fig. 3b). Precipitation  
273 variables were not very influential in PC3 (Table 1). Annual mean temperature, of the warmest and the  
274 coldest quarters, as well as ADD from July to October made a strong positive contribution to PC3, with  
275 coefficients greater than 0.90. Lower values of PC3 were obtained with increasing altitudes (Figs. 3c and  
276 S4a). CBS presences and absences were not clearly separated when plotting the values of the PCs for each  
277 citrus grid cell (Fig. S5).

278 Three PCs were extracted from the 2014 dataset, explaining 86.6% of the variability, with PC1 =  
279 40.3%, PC2 = 29.2% and PC3 = 17.1%. Precipitation variables made the greatest contributions to PC1  
280 (Table 1). The coefficient for annual precipitation was 0.87 and those for precipitation of the wettest  
281 month, wettest quarter, warmest quarter, and from October to January were greater than 0.92. Like PC2 in  
282 1950, the highest values of PC1 in 2014 were obtained in the eastern half of South Africa (Fig. 3d). In PC2,  
283 mean diurnal range made a strong positive contribution, whereas precipitation in the driest month and  
284 quarter had a strong negative influence (Table 1). The geographic representation of PC2 in 2014 was  
285 similar to that of PC1 in 1950, with the lowest values along the Indian Ocean coast (Fig. 3e). Precipitation  
286 variables had little influence on PC3 (Table 1). Annual mean temperature, ADD from July to October, as

287well as mean temperature of the warmest and coldest quarters made a strong positive contribution to PC3,  
288with coefficients greater than 0.91. Similarly to 1950, PC3 had lower values at higher altitudes (Figs. 3f  
289and S4a). CBS presences and absences were clearly discriminated when plotting the values of PC1 and  
290PC3 for each citrus grid cell (Fig. S5), with a small area of overlap corresponding to some citrus areas in  
291the Eastern Cape, Western Cape, the North West and Northern Cape (Fig. S6).

292

293Model fit and evaluation

294

295 In 1950, six climatic variables were selected with  $|r| \leq 0.7$ : maximum temperature of the warmest  
296month, minimum temperature of the coldest month, mean temperature of the driest quarter, ADD from July  
297to October, annual precipitation, and precipitation of the coldest quarter (Fig. S4). The model that included  
298the maximum temperature of the warmest month and annual precipitation showed the lowest WAIC with a  
299value of 177.51 (Table S2). When a geostatistical term was included in this model, the WAIC was reduced  
300to 126.14 (Table 2). Both climatic variables had positive estimates of their parameters. When PCs for 1950  
301were considered, the model retaining all three PCs had the lowest WAIC of 198.19 (Table S2). When a  
302geostatistical component was included in this model, the WAIC was reduced to 131.26 (Table 2) and all  
303three PCs had positive estimates of their parameters.

304 In 2014, the same six climatic variables with  $|r| \leq 0.7$  were selected (Fig. S4). The model that included  
305the maximum temperature of the warmest month, precipitation of the coldest quarter, and ADD from July  
306to October had the lowest WAIC with a value of 49.57 (Table S2). In this model, the maximum  
307temperature of the warmest month and precipitation of the coldest quarter had negative estimates of their  
308parameters, whereas that of ADD from July to October was positive (Table 2). When PCs were included,  
309the lowest WAIC of 100.70 was obtained with the model retaining only PC1 and PC3, but similar to the  
310WAIC of 101.98 with the model including all three PCs (Table S2). Both PC1 and PC3 had positive  
311estimates of their parameters (Table 2). In 2014 it was not possible to include the geostatistical term in the  
312models due to the fact that CBS presences and absences were completely separated on the map.

313 Similar predictive distributions were obtained with the models for 1950 including the maximum  
314temperature of the warmest month and annual precipitation or three PCs (Figs. 4ac). The highest  
315probabilities were obtained along the coast of Kwazulu-Natal and part of the Eastern Cape, as well as in  
316inland areas of Mpumalanga and Limpopo, with values of up to 0.93. The standard deviation associated  
317with the predictive distributions of these models was lower than 0.168, with the highest uncertainty in the

318 areas of higher probability (Figs. 4bd). The predictive distribution of the model with the two climatic  
319 variables and a geostatistical term was similar to those of the previous two models, but with a much higher  
320 probability in Kwazulu-Natal (Fig. 4e). Larger standard deviation was associated with this model, with  
321 values of up to 0.41 around the areas of high probability and in the central regions of the country (Fig. 4f).  
322 The model including three PCs and a geostatistical term predicted larger areas with a high probability of  
323 0.99, entirely covering Kwazulu-Natal and regions in Mpumalanga, Limpopo and North West provinces  
324 (Fig. 4g). Areas of high uncertainty were also much larger with this model, with values of standard  
325 deviation up to 0.44, particularly in the eastern half of the country (Fig. 4h).

326 In 2014, similar predictive distributions were obtained with the models including PC1 and PC3 or the  
327 maximum temperature of the warmest month, precipitation of the coldest quarter, and ADD from July to  
328 October (Figs. 5ac). High probabilities up to unity were obtained in Kwazulu-Natal, Mpumalanga,  
329 Limpopo, Gauteng, parts of the Eastern Cape, North West and Free State, as well as in coastal areas in the  
330 Western Cape and Northern Cape. The standard deviation associated with these predictive distributions  
331 was lower than 0.34, with the highest uncertainty around the areas of higher probability (Figs. 5bd).

332 When the predictive distributions of the models for 1950 were evaluated against the distribution of  
333 CBS in 2014, excluding those grid cells used for model development, the highest AUC of 0.986 was  
334 obtained with the model including three PCs and a geostatistical term (Fig. 6). The model with only three  
335 PCs had an AUC of 0.929. The model including the maximum temperature of the warmest month and  
336 annual precipitation had an AUC of 0.839, which was reduced to 0.821 when a geostatistical term was  
337 incorporated.

338

### 339 Discussion

340

341 Correlative species distribution models rely on the assumption that the organism modelled is in  
342 equilibrium with its environment within the region of study. Hence, the species occurs in all suitable  
343 environmental conditions (i.e. throughout the suitable environmental space), although not necessarily  
344 occupying the geographic space completely. This assumption is often violated in the case of biological  
345 invasions, where potentially suitable habitats were not yet reached by the species because of colonization  
346 time lag and/or dispersal constraints (Barve et al. 2011; Elith and Leathwick 2009). It has been stated that  
347 CBS probably attained its full potential distribution in South Africa because the disease had many

348opportunities to invade citrus areas throughout the country (Yonow et al. 2013). CBS is certainly much  
349more widespread nowadays in South Africa than in 1950 (Fig. 1), but the assumption that CBS is in  
350equilibrium with the environmental conditions and occurs in all suitable habitats in the country is  
351questionable (EFSA 2008, 2014, 2016). In fact the movement of citrus material in South Africa was not  
352regulated until 1984, although quantitative trade data among provinces was not found (Martínez-Minaya et  
353al. 2015). Since then, internal phytosanitary barriers have been in place to impede the movement of citrus  
354plants and avoid the spread of *P. citricarpa* to other regions in the country (Fig. 1b). The presence of  
355dispersal constraints for more than three decades cannot be overlooked when interpreting CBS distribution  
356in South Africa and the resulting model outcomes.

357 Process-based models comprising the entire environmental space of the species are thought to be more  
358adequate for non-equilibrium scenarios (Dormann et al. 2012b), but they still rely on disease prevalence  
359data to interpret model outputs and define thresholds for climate suitability. For instance, Magarey et al.  
360(2015) defined a specific output threshold to be able to consider a location suitable for CBS based on the  
361values for sites with moderate disease prevalence, which was the Eastern Cape in their study. However, the  
362crucial role of an accessible area fully applies to process-based models as well. Different thresholds for  
363moderate disease prevalence might be chosen considering past, present or future disease distribution data.  
364Moreover, CBS is characterized by slow epidemic development (Kotzé 1981) and thus future impacts  
365cannot be directly inferred from its present status. Besides, process-based models for CBS are seriously  
366affected by large uncertainties due to the lack of biological and epidemiological data (EFSA 2014, 2016).

367 The consideration of true absences is also a controversial issue in species distribution models. In many  
368cases, only presence data are available and models such as Maxent are preferred, which generate random  
369pseudo-absences from an area around presence records. When the species being modelled is in its early  
370stages of invasion, presence-only models are sometimes preferred because absences may not be associated  
371with climatic unsuitability (Dupin et al. 2011). However, with pseudo-absences the accuracy of the model  
372can be overestimated and reliable absence data are considered more appropriate for model validation  
373(Vaclavik and Meentemeyer 2012). In the case of the logistic regression used here, true absences are  
374required for both model development and evaluation.

375 Disease presences in the 1950 dataset were obtained from Wager (1952), who surveyed the citrus-  
376growing areas in South Africa for CBS. Nevertheless, molecular techniques for pathogen detection were  
377not available at that time and, therefore, the possibility of missing *P. citricarpa* presences in a latent  
378asymptomatic stage cannot be excluded. The map of citrus distribution in South Africa in 1927 (Powell

3791930) had a reasonable level of detail. However, it looks as if some citrus areas in the Eastern Cape might  
380be overrepresented (Fig. 1a), potentially increasing the number of CBS absences. More recent surveys for  
381CBS in South Africa should comply with international standards (IPPC 1995, 2005, 2007), which ideally  
382reduce the risk of imperfect detections and sampling bias (Guillera-Arroita et al. 2015). Although we  
383restricted our data to only citrus areas (Anonymous 2014; Paul 2005; Yonow et al. 2013), the NLS class  
384“cultivated commercial permanent orchards” also comprises other crops (DEA 2015), which potentially  
385increases CBS presences and/or absences in the 2014 dataset. On the other hand, the NLS dataset does not  
386consider ornamental or back-yard citrus trees, thereby potentially reducing CBS presences and/or absences.  
387In any case, since no other contrasting data sources were found, we consider that our analyses were based  
388on the best information available. Further refinements of our models could be possible if more accurate  
389datasets of CBS distribution in South Africa become accessible. Likewise, recent updates of the  
390WorldClim database could also be used (Fick and Hijmans 2017).

391 Significant spatial autocorrelation of CBS distribution was detected in 1950 and 2014 (Fig. 2).  
392Furthermore, the geostatistical term was relevant in the regression models for 1950, climatic variables or  
393PCs also being included as explanatory variables (Table 2). Spatial autocorrelation occurs when disease  
394observations in different locations are not independent from each other. Dispersal barriers, spatially  
395structured gradients or intrinsic spread processes usually lead to spatial autocorrelation in species  
396distribution data (Franklin 2009). The strong spatial autocorrelation detected in CBS distribution data both  
397in 1950 and 2014 suggests that climate itself might not be the main factor limiting the spread of CBS in  
398South Africa. The natural spread of CBS through *P. citricarpa* spores is poorly understood. Under  
399laboratory conditions, conidia from inoculated citrus fruit were splashed 0.6 m high and to a distance of at  
400least 8 m by simulated wind-driven rain (Perryman et al. 2014). No information is available on the  
401distances airborne ascospores of *P. citricarpa* can spread. The drivers of CBS invasions worldwide remain  
402generally unknown, but human-assisted movement of infected citrus material is considered the most  
403important means of disease introduction and spread. Regardless of the mechanisms involved in the  
404invasion process, the presence of significant spatial aggregation indicated a higher probability of CBS  
405presence in grid cells near affected areas.

406 Ignoring spatial autocorrelation in the models can decrease the precision of parameter estimates and  
407falsely reject the null hypothesis of no effect. In addition, the selection of explanatory variables may be  
408biased towards those that are more autocorrelated, such as climatic gradients. Consequently, certain  
409variables as well as more variables in general are likely to be retained, thereby making the resulting model

410potentially misleading (Chapman 2010; Dormann et al. 2007; Franklin 2009). Previous studies with  
411CLIMEX to estimate the potential geographic distribution of CBS did not consider spatial autocorrelation  
412(Er et al. 2013; Paul et al. 2005; Yonow et al. 2013). These studies were conducted on a much broader  
413geographic scale and so consequences of ignoring spatial dependence are believed to be less problematic  
414(Franklin 2009). However, the presence of dispersal constraints like phytosanitary barriers (Fig. 1b) and  
415other local range-confining processes may have limited the area and environments accessible to CBS  
416anyway. Moreover, none of these CLIMEX studies involved formal statistical inference for parameter  
417estimation and so they are difficult to compare with our models.

418 Collinearity arises when two or more explanatory variables in a model are linearly related, which is  
419common when climatic variables are considered. With collinearity, parameter estimates may be unstable  
420with inflated standard errors and thus inference may be biased and select the wrong explanatory variables.  
421Moreover, the effects of two collinear explanatory variables cannot be separated and model extrapolation  
422may be seriously flawed (Dormann et al. 2012a). PCA is one of the most common ways to manage  
423collinearity among explanatory variables in correlative species distribution models (Dupin et al. 2011;  
424Manel et al. 2001; Kriticos et al. 2014; Petitpierre et al. 2012). In our case, the use of PCs as explanatory  
425variables in the models allowed us to integrate the contribution of a relatively large set of climatic variables  
426with serious collinearity problems (Figs. S2 and S3). Nevertheless, better model fit (i.e. lower WAIC) was  
427obtained including a threshold-based pre-selection of climatic explanatory variables with pairwise  
428correlations  $|r| \leq 0.7$  (Dormann et al. 2012a).

429 In general, the regression analyses performed in our study indicated a positive relationship between  
430CBS presence and climatic variables or PCs associated with warm temperatures and high precipitation  
431(Tables 1 and 2). Indeed, it has been stated that CBS thrives mainly in warm wet climates (Yonow et al.  
4322013), although the disease is also present in arid desert conditions (Martínez-Minaya et al. 2015). Some of  
433the previous studies with CLIMEX suggested that the potential distribution of CBS could be limited by  
434cold conditions (Paul et al. 2005; Yonow et al. 2013), although these modelling approaches and their  
435parameterization have been subject to debate (EFSA 2008, 2014). In our models, degrees accumulated  
436during the period of ascocarp formation and ascospore maturation in South Africa (i.e. July to October)  
437were positively related with CBS presence (Table 2). The empirical model by Fourie et al. (2013) predicted  
438an earlier release of *Phyllosticta* spp. ascospores with warmer winters and springs, which might be  
439associated with more favourable climate conditions for CBS establishment. However, this empirical model

440 included both *P. citricarpa* and the non-pathogenic species *P. capitalensis* Henn., which is also widely  
441 established in relatively cold regions (Wikey et al. 2013).

442 Several studies have demonstrated that models for species in the early stages of invasion are more  
443 likely to underpredict potential distribution than models in advanced stages of invasion, where the  
444 equilibrium assumption is more plausible (Dupin et al. 2011; Vaclavik and Meentemeyer 2012). In our  
445 case, relatively high accuracy was obtained with the models for the 1950 dataset, representing the early  
446 stages of CBS epidemics in South Africa. An AUC of 0.929 was obtained with the model including PCs  
447 and an AUC of 0.986 resulted when a geostatistical term was also incorporated (Fig. 6). According to the  
448 criteria put forward by Swets (1988), these AUC values are indicative of rather high accuracy. However,  
449 despite their good accuracy, none of our models were able to predict subsequent CBS invasions in citrus  
450 areas in the Eastern Cape and north of Limpopo (Fig. 4), where the disease thrives under more arid  
451 conditions (Martínez-Minaya et al. 2015).

452 Citrus areas in the north of Limpopo are considered of low pest (disease) prevalence (Anonymous  
453 2014), which implies that CBS occurs at low levels and is subjected to effective surveillance, control or  
454 eradication measures (IPPC 2005, 2007). It has been claimed that CBS has low or moderate prevalence in  
455 the Eastern Cape (Fourie et al. 2013; Magarey et al. 2015), but this region is not officially considered an  
456 area of low CBS prevalence (Anonymous 2014). Moreover, Schutte (1995) indicated that in the Eastern  
457 Cape lemons were sprayed with fungicides for CBS control. This has been confirmed by more recent  
458 reports, indicating that fungicide applications for CBS control have increased in the Eastern Cape and  
459 lemons must be frequently sprayed (Grout 2015).

460 Citrus-growing areas in the Eastern Cape are the only ones affected by CBS nowadays that were left  
461 outside the phytosanitary barrier established in 1984 (Fig. 1b). Considering the long lag phase of CBS  
462 epidemics (Kotzé 1981) and that fungicides were applied for its control in the Eastern Cape in the 1990s  
463 (Schutte 1995), it is conceivable that this region was already affected several years before, but perhaps at  
464 very low levels or still in an asymptomatic stage. Indeed, Kotzé (1981) indicated that *P. citricarpa* may be  
465 present for many years in a particular area before CBS symptoms appear. Consequently, citrus areas in the  
466 Eastern Cape might have been inadvertently considered as free of *P. citricarpa* when designing the  
467 phytosanitary barrier in 1984. Interestingly, when representing PC1 and PC3 associated with precipitation  
468 and warm temperatures in 2014, citrus areas in the Eastern Cape currently affected by CBS overlapped  
469 with some CBS-free areas in the Western Cape, Northern Cape and North West provinces (Fig. S6).

470Although the models for 2014 did not incorporate a geostatistical term, those including climatic variables  
471or PCs displayed relatively high probabilities of CBS presence in these particular areas (Fig. 5). Therefore,  
472intensive surveys would be recommended to keep them free from disease.

473 Although climate has been advocated as the main factor limiting the establishment and spread of CBS  
474into new areas (Magarey et al. 2015; Paul et al. 2005; Yonow et al. 2013), our study indicates that spatial  
475proximity to affected areas is also relevant in the geographic distribution of the disease in South Africa.  
476Indeed, some historical evidence illustrated that too much hope had been pinned on climate as a limiting  
477factor for CBS (Kotzé 1981). In his detailed study, Wager (1952) indicated that CBS was first reported in  
478South Africa by Doidge (1929) in a relatively cool mist-belt area with high rainfall. As at that time it was  
479assumed that CBS required this type of conditions, no concern was therefore felt for its possible spread to  
480other parts of South Africa (Wager 1952). However, from 1940 to 1950 the disease spread to neighbouring  
481citrus regions under much drier conditions. Based on this, Wager (1952) concluded that the old concept of  
482CBS requiring cool, moist, or mist-belt conditions for its development was wrong. McOnie (1964c)  
483surveyed the citrus areas in the Eastern Cape and concluded that *P. citricarpa* was absent due to  
484unfavourable climatic conditions. However, the pathogen was later reported in the Eastern Cape and  
485fungicide sprays are currently applied for CBS control (Grout 2015; Schutte 1995). In Zimbabwe,  
486Whiteside (1965) stated that CBS may not become really serious under local climatic conditions.  
487Nevertheless, the disease reached epidemic proportions in 1978 (Kotzé 1981). More recently, Guarnaccia  
488et al. (2017) reported, for the first time, the presence of *P. citricarpa* in the Mediterranean Basin, under  
489dry-summer climate conditions. The future will determine whether current models for the potential  
490geographical distribution of CBS can stand the test of time.

491

#### 492**Acknowledgements**

493Authors would like to thank Iosu Paradinas from UV and Xavier Barber from MHU for support with the  
494use of R and R-INLA, and the INLA-project team, in particular, Elias T. Krainski from UF Paraná, for  
495their prompt support with technical aspects of the usage of R-INLA, and to V. Monzó (MON Topografía y  
496Cartografía) for digitizing and georeferencing maps.

497

#### 498**Compliance with Ethical Standards**

499Disclosure of potential conflicts of interest: DC and AL-Q are supported by MECD ERDF grant  
500MTM2016-77501-P. JM-M is supported by GVA grant VALi+d ACIF/2016/455.

501

502**References**

- 503 Anonymous (1984). R.110 Agricultural Pest Act, 1983 (Act 36 of 1983). Control measures. *Government*  
504 *Gazette, 9047*, 6-11.
- 505 Anonymous (2002). R.831 Agricultural Pest Act, 1983 (Act 36 of 1983). Control measures: Amendment.  
506 *Government Gazette, 23517*, 15-17.
- 507 Anonymous (2014). R.442 Agricultural Pest Act, 1983 (Act 36 of 1983). Control measures: Amendment.  
508 *Government Gazette, 37702*, 4-11.
- 509 Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberon, J., &  
510 Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species  
511 distribution modeling. *Ecological Modelling, 222*, 1810-1819.
- 512 Chapman, D. S. (2010). Weak climatic associations among British plant distributions. *Global Ecology and*  
513 *Biogeography, 19*, 831-841.
- 514 Chatfield, C., & Collins, A. J. (2013). *Introduction to multivariate analysis*. Berlin: Springer.
- 515 DEA, Department of Environmental Affairs South Africa (2015). South African national land-cover  
516 dataset 2013-2014, Geoterraimage. <http://egis.environment.gov.za>, accessed on 8 February 2016.
- 517 Doidge, E. M. (1929) Some diseases of citrus prevalent in South Africa. *South African Journal of Science*,  
518 *26*, 320-325.
- 519 Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., García-Márquez, J. R., Gruber, B.,  
520 Lafourcade, B., Leitao, P. J., Muenkemueller, T., McClean, C., Osborne, P. E., Reineking, B.,  
521 Schroeder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2012a). Collinearity: a review of  
522 methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*, 27-46.
- 523 Dormann, C. F., McPherson, J. M., Araujo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel,  
524 A., Jetz, W., Kissling, W. D., Kuehn, I., Ohlemueller, R., Peres-Neto, P. R., Reineking, B., Schroeder,  
525 B., Schurr, F. M., & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of  
526 species distributional data: a review. *Ecography, 30*, 609-628.
- 527 Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X.,  
528 Roemermann, C., Schroeder, B., & Singer, A. (2012b). Correlation and process in species distribution  
529 models: bridging a dichotomy. *Journal of Biogeography, 39*, 2119-2131.

530Dupin, M., Reynaud, P., Jarošik, V., Baker, R., Brunel, S., Eyre, D., Pergl, J., & Makowski, D. (2011).  
531 Effects of the training dataset characteristics on the performance of nine species distribution models:  
532 Application to *Diabrotica virgifera virgifera*. *Plos One*, 6, e20957.

533EFSA, European Food Safety Authority. (2008). Pest risk assessment and additional evidence provided by  
534 South Africa on *Guignardia citricarpa* Kiely, citrus black spot fungus–CBS. Scientific Opinion of the  
535 PLH Panel, European Food Safety Authority. *EFSA Journal*, 925, 1-108.

536EFSA, European Food Safety Authority. (2014). Scientific opinion on the risk of *Phyllosticta citricarpa*  
537 (*Guignardia citricarpa*) for the EU territory with identification and evaluation of risk reduction options.  
538 *EFSA Journal*, 12, 3557.

539EFSA, European Food Safety Authority. (2016). Evaluation of new scientific information on *Phyllosticta*  
540 *citricarpa* in relation to the EFSA PLH Panel (2014). Scientific Opinion on the plant health risk to the  
541 EU. *EFSA Journal*, 14, 4513.

542Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction  
543 across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.

544Elith, J., Simpson, J., Hirsch, M., & Burgman, M. A. (2013). Taxonomic uncertainty and decision making  
545 for biosecurity: spatial models for myrtle/guava rust. *Australasian Plant Pathology*, 42: 43-51.

546Er, H. L., Roberts, P. D., Marois, J. J., & van Bruggen, A. H. C. (2013). Potential distribution of citrus  
547 black spot in the United States based on climatic conditions. *European Journal of Plant Pathology*,  
548 137, 635-647.

549Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global  
550 land areas. *International Journal of Climatology*, DOI 10.1002/joc.5086.

551Fourie, P. H., Schutte, G. C., Serfontein, S., & Swart, S. H. (2013). Modeling the effect of temperature and  
552 wetness on *Guignardia pseudothecium* maturation and ascospore release in citrus orchards.  
553 *Phytopathology*, 103, 281-292.

554Franklin, J. (2009). *Mapping species distributions. Spatial inference and prediction*. New York: Cambridge  
555 University Press.

556Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian  
557 models. *Statistics and Computing*, 24, 997-1016.

558Grout, T. G. (2015). The status of citrus IPM in South Africa. *Acta Horticulturae*, 1065, 1091-1095.

559Guarnaccia, V., Groenewald, J. Z., Li, H., Glienke, C., Carstens, E., Hattingh, V., Fourie, P. H., & Crous,  
560 P. W. (2017). First report of *Phyllosticta citricarpa* and description of two new species, *P.*  
561 *paracapitalensis* and *P. paracitricarpa*, from citrus in Europe. *Studies in Mycology*, 87, 161-185.

562Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M.  
563 A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching  
564 data and models to applications. *Global Ecology and Biogeography*, 24, 276-292.

565Hendricks, K. E., Christman, M., & Roberts, P. D. (2017). Spatial and temporal patterns of commercial  
566 citrus trees affected by *Phyllosticta citricarpa* in Florida. *Scientific Reports*, 7, 1641.

567Hijmans, R. J. (2014). raster: geographic data analysis and modeling. R package version 2.2-31.  
568 <http://CRAN.R-project.org/package=raster>.

569Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution  
570 interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-  
571 1978.

572IPPC, International Plant Protection Convention. (1995). *Requirements for the establishment of pest free*  
573 *areas. International Standards for Phytosanitary Measures, ISPM 4*. Rome: IPPC.

574IPPC, International Plant Protection Convention. (2005). *Requirements for the establishment of areas of*  
575 *low pest prevalence. International Standards for Phytosanitary Measures, ISPM 22*. Rome: IPPC.

576IPPC, International Plant Protection Convention. (2007). *Recognition of pest free areas and areas of low*  
577 *pest prevalence. International Standards for Phytosanitary Measures, ISPM 29*. Rome: IPPC.

578Johnson, R. A., & Wichern, D. W. (2002) *Applied multivariate statistical analysis, 5<sup>th</sup>ed*. New Jersey:  
579 Prentice Hall.

580Kotzé, J. M. (1981). Epidemiology and control of citrus black spot in South-Africa. *Plant Disease*, 65, 945-  
581 950.

582Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand.  
583 *Journal of the Southern African Institute of Mining and Metallurgy*, 52, 119-139.

584Kriticos, D. J., Jarošik, V., & Ota, N. (2014). Extending the suite of BIOCLIM variables: a proposed  
585 registry system and case study using principal components analysis. *Methods in Ecology and*  
586 *Evolution*, 5, 956-960.

587Latimer, A. M., Wu, S. S., Gelfand, A. E., & Silander, J. A. (2006). Building statistical models to analyze  
588 species distributions. *Ecological Applications*, 16, 33-50.

589Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*,  
590 63, 19

591Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian  
592 Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal*  
593 *Statistical Society: Series B*, 73, 423-498.

594Magarey, R., Chanelli, S., & Holtz, T. (2011). *Validation study and risk assessment: Guignardia*  
595 *citricarpa*, (citrus black spot). *USDA-APHIS-PPQ-CPHST-PERAL /NCSU*.

596Magarey, R. D., Hong, S. C., Fourie, P. H., Christie, D. N., Miles, A. K., Schutte, G. C., & Gottwald, T. R.  
597 (2015). Prediction of *Phyllosticta citricarpa* using an hourly infection model and validation with  
598 prevalence data from South Africa and Australia. *Crop Protection*, 75, 104-114.

599Magarey R, Sutton T, & Thayer C (2005). A simple generic infection model for foliar fungal plant  
600 pathogens. *Phytopathology*, 95, 92-100.

601Makowski, D., Vicent, A., Pautasso, M., Stancanelli, G., & Rafoss, T. (2014). Comparison of statistical  
602 models in a meta-analysis of fungicide treatments for the control of citrus black spot caused by  
603 *Phyllosticta citricarpa*. *European Journal of Plant Pathology*, 139, 79-94.

604Manel, S., Williams, H. C., & Ormerod, S. J. (2001). Evaluating presence-absence models in ecology: the  
605 need to account for prevalence. *Journal of Applied Ecology*, 38, 921-931.

606Martínez-Minaya, J., Conesa, D., López-Quílez, A., & Vicent, A. (2015). Climatic distribution of citrus  
607 black spot caused by *Phyllosticta citricarpa*. A historical analysis of disease spread in South Africa.  
608 *European Journal of Plant Pathology*, 143, 69-83.

609McOnie, K. C. (1964a). The latent occurrence in citrus and other hosts of *Guignardia* easily confused with  
610 *G. citricarpa*, the citrus black spot pathogen. *Phytopathology*, 54, 40-43.

611McOnie, K. C. (1964b). Orchard development and discharge of ascospores of *Guignardia citricarpa* and  
612 onset of infection in relation to control of citrus black spot. *Phytopathology*, 54, 1448-1454.

613McOnie, K. C. (1964c). Apparent absence of *Guignardia citricarpa* Kiely from localities where citrus  
614 black spot is absent. *South African Journal of Agricultural Science*, 7, 347-354.

615Meentemeyer, R. K., Anacker, B. L., Mark, W., & Rizzo, D. M. (2008). Early detection of emerging forest  
616 disease using dispersal estimation and ecological niche modeling. *Ecological Applications*, 18, 377-  
617 390.

618 Muñoz, F., Pennino, M. G., Conesa, D., López-Quílez, A., & Bellido, J. M. (2013). Estimation and  
619 prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. *Stochastic*  
620 *Environmental Research and Risk Assessment*, *27*, 1171-1180.

621 Narouei-Khandan, H. A., Harmon, C. L., Harmon, P., Olmstead, J., Zelenev, V. V., van der Werf W.,  
622 Worner S. P., Senay S. D., & van Bruggen, A. H. C. (2017). Potential global and regional geographic  
623 distribution of *Phomopsis vaccinii* on *Vaccinium* species projected by two species distribution models.  
624 *European Journal of Plant Pathology*, *148*, 919-930.

625 Paul, I. (2005). *Modelling the distribution of citrus black spot caused by Guignardia citricarpa Kiely*. PhD  
626 Thesis. Pretoria: University of Pretoria.

627 Paul, I., van Jaarsveld, A.S., Korsten, L., & Hattingh, V. (2005). The potential global geographical  
628 distribution of citrus black spot caused by *Guignardia citricarpa* Kiely: likelihood of disease  
629 establishment in the European Union. *Crop Protection*, *24*, 297-308.

630 Perryman, S. A. M., Clark, S. J., & West, J. S. (2014). Splash dispersal of *Phyllosticta citricarpa* conidia  
631 from infected citrus fruit. *Scientific Reports*, *4*, 6568.

632 Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., & Guisan, A. (2012). Climatic niche  
633 shifts are rare among terrestrial plant invaders. *Science*, *335*, 1344-1348.

634 Plant, R. E. (2012). *Spatial data analysis in ecology and agriculture using R*. Boca Raton: CRC Press.

635 Powell, H. C. (1930). *The culture of the orange and allied fruits. South African agricultural series No. 8*.  
636 Johannesburg: Central News Agency.

637 R Core Team (2016). R: A language and environment for statistical computing. R Foundation for  
638 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

639 Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Mueller, M. (2011). pROC: an  
640 open-source package for R and S plus to analyze and compare ROC curves. *BMC Bioinformatics*, *12*,  
641 77.

642 Rue, H., & Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Boca Raton: CRC  
643 Press.

644 Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by  
645 using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B*, *71*,  
646 319-392.

647 Schutte, G. C. (1995). *Evaluation of control strategies for citrus black spot in Southern Africa*. PhD Thesis.  
648 Pretoria: University of Pretoria.

649Spósito, M. B., Amorim, L., Bassanezi, R. B., & Hau, B. (2008). Spatial pattern of black spot incidence  
650 within citrus trees related to disease severity and pathogen dispersal. *Plant Pathology*, *57*, 103-108.

651Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285-1293.

652Tran, N. T., Miles, A. K., Dietzgen, R. G., Dewdney, M. M., Zhang, K., Rollins, J. A., & Drenth, A.  
653 (2017). Sexual reproduction in the citrus black spot pathogen, *Phyllosticta citricarpa*. *Phytopathology*,  
654 *107*, 732-739.

655Vaclavik, T., & Meentemeyer, R. K. (2012). Equilibrium or not? Modelling potential distribution of  
656 invasive species in different stages of invasion. *Diversity and Distributions*, *18*, 73-83.

657Venette, R. C., Kriticos, D. J., Magarey, R. D., Koch, F. H., Baker, R. H. A., Worner, S. P., Raboteaux, N.  
658 N. G., McKenney, D. W., Dobesberger, E. J., Yemshanov, D., De Barro, P. J., Hutchison, W. D.,  
659 Fowler, G., Kalaris, T. M., & Pedlar, J. (2010). Pest risk maps for invasive alien species: A roadmap for  
660 improvement. *Bioscience*, *60*, 349-362.

661Wager, V. A. (1952). The black spot disease of citrus in South Africa. *Science Bulletin of the Department*  
662 *of Agriculture of the Union of South Africa*, *303*, 1-52.

663Wang, N. Y., Zhang, K., Huguet-Tapia, J. C., Rollins, J. A., & Dewdney, M. M. (2016). Mating type and  
664 simple sequence repeat markers indicate a clonal population of *Phyllosticta citricarpa* in Florida.  
665 *Phytopathology*, *106*, 1300-1310.

666Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information  
667 criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571-3594.

668Whiteside, J. O. (1965). Black spot disease in Rhodesia. *Rhodesian Agricultural Journal*, *63*, 87-91.

669Wikee, S., Lombard, L., Crous, P. W., Nakashima, C., Motohashi, K., Chukeatirote, E., Alias, S. A.,  
670 McKenzie, E. H. C., & Hyde, K. D. (2013). *Phyllosticta capitalensis*, a widespread endophyte of plants.  
671 *Fungal Diversity*, *60*, 91-105.

672Yonow, T., Hattingh, V., & de Villiers, M. (2013). CLIMEX modelling of the potential global distribution  
673 of the citrus black spot disease caused by *Guignardia citricarpa* and the risk posed to Europe. *Crop*  
674 *Protection*, *44*, 18-28.

675**Figure 1:** Citrus-growing areas and distribution of citrus black spot (CBS) in South Africa in **a** 1950 and **b**  
6762014, with lines indicating the prohibition boundary for the east-west movement of citrus plants in 1984  
677(dashed line) and 2002 (solid line) (Anonymous 1984, 2002; DEA, 2015; Martínez-Minaya et al. 2015;  
678Paul 2005; Powell, 1930; Wager 1952; Yonow et al. 2013). Data for Lesotho and Swaziland were not  
679available. **c** Solid lines indicate province boundaries.

680**Figure 2:** Moran's I (a) and Geary's C (b) values for contiguity and at increasing distances, with orange  
681lines for 1950 and red lines for 2014.

682**Figure 3:** Geographic representation of the rotated principal components PC1 (a), PC2 (b), PC3 (c) for  
6831950 and PC1 (d), PC2 (e) and PC3 (f) for 2014.

684**Figure 4:** Mean (red) and standard deviation (blue) of the predictive posterior distribution for the  
685probability of citrus black spot presence with the best models of 1950 including climatic variables (a,b),  
686principal components (c,d), climatic variables + geostatistical term (e, f) and principal components +  
687geostatistical term (g, h).

688**Figure 5:** Mean (red) and standard deviation (blue) of the predictive posterior distribution for the  
689probability of citrus black spot presence with the best models of 2014 including climatic variables (a,b) or  
690principal components (c,d).

691**Figure 6:** Receiver operating characteristic (ROC) curves and area under the curve (AUC) obtained with  
692the 2014 validation dataset with the best models for the probability of citrus black spot presence in 1950  
693including climatic variables (a), climatic variables + geostatistical term (b), principal components (c) and  
694principal components + geostatistical term (d).

695**Figure S1:** Validation dataset with citrus black spot (CBS) presences ( $n = 385$ ) and absences ( $n = 259$ ) in  
6962014, excluding those grid cells used for model development in 1950.

697**Figure S2:** Correlation matrix for the climatic variables of the 1950 dataset of citrus black spot distribution  
698in South Africa.

699**Figure S3:** Correlation matrix for the climatic variables of the 2014 dataset of citrus black spot distribution  
700in South Africa.

701**Figure S4:** Maps of **a** altitude; **b** maximum temperature of the warmest month (BIO<sub>5</sub>); **c** minimum  
702temperature of the coldest month (BIO<sub>6</sub>); **d** mean temperature of the driest quarter (BIO<sub>9</sub>); **e** annual  
703precipitation (BIO<sub>12</sub>); **f** precipitation of the coldest quarter (BIO<sub>19</sub>); and **g** accumulated degrees (ADD) from  
704July to October with  $T_{base} = 10^{\circ}\text{C}$  for South Africa obtained from the WorldClim database (Hijmans et al.  
7052005).

706**Figure S5:** Scatterplots of the principal components for 1950 (**a,b,c**) and 2014 (**d,e,f**). Red dots are grid  
707cells with citrus black spot (CBS) presence and green dots denote those with CBS absence.

708**Figure S6:** Scatterplot of the principal components PC1 and PC3 in 2014 with their corresponding 95%  
709confidence ellipses (**a**), and a map representing the grid cells within the area of overlap of the two ellipses  
710(**b**). Red dots are grid cells with citrus black spot (CBS) presence and green dots denote those with CBS  
711absence.

712**Table 1:** Climatic variables (BIO) and three linear combinations (PC) extracted with principal component analysis (PCA) in the 1950 and  
7132014 datasets and their explained variability.

714

Climatic variables <sup>1</sup>		1950			2014		
		PC1	PC2	PC3	PC1	PC2	PC3
BIO <sub>1</sub>	Annual mean temperature	0.180	0	0.976	0.133	0.133	0.978
BIO <sub>2</sub>	Mean diurnal range (mean of monthly (max temp – min temp))	0.932	-0.203	0	-0.307	0.900	0
BIO <sub>3</sub>	Isothermality (BIO <sub>2</sub> /BIO <sub>7</sub> )	0	0.433	0	0.661	0.106	0.328
BIO <sub>4</sub>	Temperature seasonality (standard deviation * 100)	0.887	-0.313	-0.150	-0.594	0.667	-0.258
BIO <sub>5</sub>	Max temperature of warmest month	0.558	-0.361	0.695	-0.526	0.453	0.651
BIO <sub>6</sub>	Min temperature of coldest month	-0.666	0	0.726	0.139	-0.638	0.716
BIO <sub>8</sub>	Mean temperature of wettest quarter	0.447	0.408	0.647	0.451	0.367	0.741
BIO <sub>9</sub>	Mean temperature of driest quarter	-0.323	-0.448	0.518	-0.635	-0.533	0
BIO <sub>10</sub>	Mean temperature of warmest quarter	0.365	-0.106	0.909	-0.159	0.298	0.915
BIO <sub>11</sub>	Mean temperature of coldest quarter	-0.236	0.136	0.958	0.273	-0.172	0.943
BIO <sub>12</sub>	Annual precipitation	-0.287	0.896	-0.137	0.873	-0.400	0
BIO <sub>13</sub>	Precipitation of wettest month	0.156	0.925	0	0.927	0	0
BIO <sub>14</sub>	Precipitation of driest month	-0.876	-0.242	-0.230	0	-0.887	-0.184
BIO <sub>15</sub>	Precipitation seasonality (Coefficient of variation)	0.719	0.522	0.250	0.408	0.664	0.287
BIO <sub>16</sub>	Precipitation of wettest quarter	0.129	0.936	0	0.930	-0.130	0
BIO <sub>17</sub>	Precipitation of driest quarter	-0.879	-0.238	-0.259	0	-0.889	-0.216
BIO <sub>18</sub>	Precipitation of warmest quarter	0.166	0.970	0	0.929	0.102	0.302
BIO <sub>19</sub>	Precipitation of coldest quarter	-0.606	-0.434	-0.289	-0.384	-0.599	-0.511
AP	Precipitation from October to January	0.125	0.974	0	0.948	0	0.231
ADD	Accumulated degrees from July to October with $T_{base} = 10^{\circ}\text{C}$	0.177	0.200	0.928	0.312	0.104	0.912
% variability:		37.4	25.9	22.1	40.3	29.2	17.1

715 <sup>1</sup>Temperature variables in °C and precipitation variables in mm. The variable temperature annual range (BIO<sub>7</sub>) was not included in the PCA because it  
716 is a linear combination of BIO<sub>5</sub> and BIO<sub>6</sub>.

717**Table 2:** Best models for 1950 and 2014 with climatic variables (BIO), principal components (PC) and  
718geostatistical term ( $W$ ).

719

Models <sup>1</sup>			WAIC <sup>2</sup>
<b>1950</b>			
1 + BIO <sub>5</sub> + BIO <sub>12</sub>			177.51
	Mean	Sd	
Intercept	-26.593	5.401	
BIO <sub>5</sub>	0.478	0.146	
BIO <sub>12</sub>	0.012	0.002	
1 + PC <sub>1</sub> + PC <sub>2</sub> + PC <sub>3</sub>			198.19
	Mean	Sd	
Intercept	-4.481	0.391	
PC <sub>1</sub>	-0.640	0.240	
PC <sub>2</sub>	1.325	0.231	
PC <sub>3</sub>	0.515	0.209	
1 + BIO <sub>5</sub> + BIO <sub>12</sub> + $W$			126.14
	Mean	Sd	
Intercept	-95.237	35.005	
BIO <sub>5</sub>	2.151	0.859	
BIO <sub>12</sub>	0.031	0.009	
1 + PC <sub>1</sub> + PC <sub>2</sub> + PC <sub>3</sub> + $W$			131.26
	Mean	Sd	
Intercept	-5.933	4.137	
PC <sub>1</sub>	2.037	2.732	
PC <sub>2</sub>	5.539	2.588	
PC <sub>3</sub>	4.620	2.257	
<b>2014</b>			
1 + BIO <sub>5</sub> + BIO <sub>19</sub> + ADD			49.57
	Mean	Sd	
Intercept	48.910	11.716	
BIO <sub>5</sub>	-2.886	0.616	
BIO <sub>19</sub>	-0.126	0.028	
ADD	0.770	0.159	
1 + PC <sub>1</sub> + PC <sub>3</sub>			100.70
	Mean	Sd	
Intercept	7.934	1.211	
PC <sub>1</sub>	9.145	1.259	
PC <sub>2</sub>	8.242	1.156	

720

721

722

723

724

725

726

727

<sup>1</sup>Maximum temperature of warmest month (BIO<sub>5</sub>), minimum temperature of coldest month (BIO<sub>6</sub>), mean temperature of driest quarter (BIO<sub>9</sub>), accumulated degrees (ADD) from July to October with  $T_{base} = 10^{\circ}\text{C}$ , annual precipitation (BIO<sub>12</sub>) and precipitation of coldest quarter (BIO<sub>19</sub>).

<sup>2</sup>Watanabe Akaike Information Criterion. Lower values of WAIC reflect a better model fit balanced with model complexity.

728**Table S2:** Best models for 1950 and 2014 with climatic variables (BIO), principal components (PC) and  
729geostatistical term ( $W$ ).

730

Models <sup>1</sup>	WAIC <sup>2</sup>		
1950			
Climatic + spatial	1 + BIO <sub>5</sub> + BIO <sub>12</sub> + $W$	126.14	
	1 + BIO <sub>5</sub> + BIO <sub>12</sub>	177.51	
	1 + BIO <sub>5</sub> + BIO <sub>12</sub> + BIO <sub>9</sub> + BIO <sub>19</sub> + ADD	178.07	
	1 + BIO <sub>12</sub> + BIO <sub>9</sub> + BIO <sub>19</sub> + ADD	178.26	
	1 + BIO <sub>5</sub> + BIO <sub>12</sub> + BIO <sub>19</sub>	178.85	
	1 + BIO <sub>6</sub> + BIO <sub>12</sub> + BIO <sub>9</sub> + BIO <sub>19</sub> + ADD	178.93	
	1 + BIO <sub>5</sub> + BIO <sub>12</sub> + BIO <sub>9</sub>	178.95	
	1 + BIO <sub>5</sub> + BIO <sub>12</sub> + ADD	179.01	
	1 + BIO <sub>5</sub> + BIO <sub>6</sub> + BIO <sub>12</sub> + BIO <sub>9</sub> + BIO <sub>19</sub> + ADD	179.57	
	1 + BIO <sub>5</sub> + BIO <sub>6</sub> + BIO <sub>12</sub>	179.61	
	1 + BIO <sub>12</sub> + ADD	179.73	
	PC + spatial	1 + PC <sub>1</sub> + PC <sub>2</sub> + PC <sub>3</sub> + $W$	131.26
		1 + PC <sub>1</sub> + PC <sub>2</sub> + PC <sub>3</sub>	198.19
1 + PC <sub>1</sub> + PC <sub>2</sub>		202.2	
1 + PC <sub>2</sub> + PC <sub>3</sub>		202.44	
1 + PC <sub>2</sub>		208.52	
1 + PC <sub>1</sub> + PC <sub>3</sub>		235.49	
1 + PC <sub>1</sub>		236.56	
1 + PC <sub>3</sub>		244.15	
1		245.06	
2014			
Climatic		1 + BIO <sub>5</sub> + BIO <sub>19</sub> + ADD	49.57
		1 + BIO <sub>5</sub> + BIO <sub>6</sub> + BIO <sub>19</sub> + ADD	52.23
		1 + BIO <sub>6</sub> + BIO <sub>9</sub> + BIO <sub>12</sub> + BIO <sub>19</sub> + ADD	63.77
	1 + BIO <sub>6</sub> + BIO <sub>12</sub> + BIO <sub>19</sub> + ADD	65.68	
	1 + BIO <sub>5</sub> + BIO <sub>6</sub> + BIO <sub>19</sub>	70.11	
	1 + BIO <sub>6</sub> + BIO <sub>12</sub> + BIO <sub>19</sub>	70.52	
	1 + BIO <sub>6</sub> + BIO <sub>9</sub> + BIO <sub>12</sub> + BIO <sub>19</sub>	72.1	
	1 + BIO <sub>5</sub> + BIO <sub>6</sub> + BIO <sub>9</sub> + ADD	72.79	
	1 + BIO <sub>5</sub> + BIO <sub>6</sub> + BIO <sub>9</sub> + BIO <sub>12</sub> + ADD	73.67	
	1 + BIO <sub>5</sub> + BIO <sub>6</sub> + ADD	77.68	
	PC	1 + PC <sub>1</sub> + PC <sub>3</sub>	100.7
		1 + PC <sub>1</sub> + PC <sub>2</sub> + PC <sub>3</sub>	101.98
		1 + PC <sub>1</sub> + PC <sub>2</sub>	439.13
1 + PC <sub>1</sub>		520.41	
1 + PC <sub>2</sub> + PC <sub>3</sub>		892.93	
1 + PC <sub>3</sub>		946.62	
1 + PC <sub>2</sub>		1144.74	
1		1192.49	

731

732

733

734

735

736

<sup>1</sup>Maximum temperature of warmest month (BIO<sub>5</sub>), minimum temperature of coldest month (BIO<sub>6</sub>), mean temperature of driest quarter (BIO<sub>9</sub>), accumulated degrees (ADD) from July to October with  $T_{base} = 10^{\circ}\text{C}$ , annual precipitation (BIO<sub>12</sub>) and precipitation of coldest quarter (BIO<sub>19</sub>).

<sup>2</sup>Watanabe Akaike Information Criterion. Lower values of WAIC reflect a better model fit balanced with model complexity.