^aNatural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado, USA.

^bU.S. Geological Survey Earth Resources Observation and Science (EROS) Center/ North Central Climate Science Center, Fort Collins, Colorado, USA.

^cCollege of Agriculture and Veterinary, Plant Sciences Department, Addis Ababa University, Ethiopia.

^dCollege of Agriculture and Veterinary, Plant Sciences Department, Ambo University, Ethiopia.

E-mail address: Tewodros.Wakie@colostate.edu

Abstract

A number of studies have reported the presence of wheat septoria leaf blotch (*Septoria tritici*; SLB) disease in Ethiopia. However, the environmental factors associated with SLB disease, and areas under risk of SLB disease, have not been studied. Here, we tested the hypothesis that environmental variables can adequately explain observed SLB disease severity levels in West Shewa, central Ethiopia. Specifically, we identified 50 environmental variables and assessed their relationships with SLB disease severity. Geographically referenced disease severity data were obtained from the field, and linear regression and Boosted Regression Trees (BRT) modeling approaches were used for developing spatial models. Moderate-resolution imaging spectroradiometer (MODIS) derived vegetation indices and land surface temperature (LST) variables highly influenced SLB model predictions. Soil and topographic variables did not sufficiently explain observed SLB disease severity variation in this study. Our results show that wheat growing areas in central Ethiopia, including highly productive districts, are at risk of SLB disease. The study demonstrates the integration of field data with modeling approaches such as BRT for predicting the spatial patterns of severity of a pathogenic wheat disease in central Ethiopia. Our results can aid Ethiopia's wheat disease monitoring efforts, while our methods can be replicated for testing related hypotheses elsewhere.

Key words: Boosted regression trees, disease severity, MODIS, pathogen risk, vegetation indices, wheat production

1. Introduction

Wheat is the third most widely produced cereal crop in the world and the fourth most important staple crop in Ethiopia. As the human population rises and moves into the urban centers, wheat demand in Ethiopia is expected to increase sharply (Macauley, 2015). Over the past decade wheat productivity in Ethiopia has moderately increased (Dercon and Hill, 2009; Taffesse et al., 2012). However, in order to achieve food self-sufficiency and food security, Ethiopia needs to increase its wheat productivity at a higher rate, while minimizing yield loss from diseases and pests. Our goal was to identify wheat growing areas in West Shewa (Ethiopia) that are vulnerable to attacks from the pathogenic fungal species called *Zymoseptoria tritici* (Desm).

In Ethiopia, wheat is mainly grown in the highlands during the main crop season (locally called *meher*). Depending on the availability of the rains and the crop varieties used, seed sowing occurs in the *meher* season from June to July, while harvesting takes place from October to January. A minor, less reliable and unpredictable crop season named *belg* also occurs from March to May. More than 90% of crop yield in Ethiopia is produced during *meher* season (CSA, 2015); thus, the focus of this study is the *meher* season. Rain-fed farming in Ethiopia is vulnerable to natural drought, particularly during the El Niño and La Niña years (Broad and Agrawala, 2000; Lyon, 2014; WFP, 2015). Fungal diseases are another threat that pose substantial risk to Ethiopian wheat production, often causing up to 50% yield loss on susceptible varieties (Abebe et al., 2015; Ponomarenko et al., 2011; Takele et al., 2015a). World-wide annual wheat yield loss from only two pathogenic genera, *Septoria* and *Stagonospora*, are estimated in the billions of dollars (Eyal, 1999; Fones and Gurr, 2015; Ponomarenko et al., 2011).

Zymoseptoria tritici (synonym *Mycosphaerella graminicola* or *Septoria tritici*) is the fungal pathogen that causes the wheat disease, referred to as *Septoria tritici* blotch (STB) or septoria leaf blotch (SLB)¹. *Mycosphaerella graminicola* was rated as one of the top 10 economically important fungal pathogens in the world (Dean et al., 2012). A number of studies have reported the occurrence of *M. graminicola* and SLB in Ethiopian wheat farms (Arraiano and Brown, 2006; Beyene et al., 1990; Takele et al., 2015b; Tanner and Hulluka, 1991). *M. graminicola* is primarily disseminated through airborne ascospores, and pycnidiospores that linger on crop residues. Spores grow on leaves and inoculate the host via stomatal openings. The life cycle of *M. graminicola* has been described using distinct stages and phases (O'Driscoll et al., 2014; Ponomarenko et al., 2011). These include hyphal growth on leaf surface, host penetration via stomata, intercellular hyphal growth, necrotrophic growth, and formation of asexual (conidia, pycnidia) and sexual (pseudothecia) fruiting bodies (Dean et al., 2012; Eyal, 1999; Ponomarenko et al., 2011). The necrotrophic stage is characterized by formation of lesions on infected leaves and stems. Conidia spreads upward on the same plant and to nearby plants, with rain-splash aiding the secondary infection and dissemination processes. Under the right environmental conditions, spread via conidia may occur several times and the infection may quickly reach epidemic levels.

Wet and cool environmental conditions are known to favor *M. graminicola* infection and spread. Wet conditions lasting longer than 12 hours and temperatures ranging from 7°C to 22°C generally create suitable conditions for *M. graminicola* invasion (Magboul et al., 1992; Te Beest et al., 2009). Soil conditions such as high soil clay content and application of nitrogen fertilizer may also influence SLB occurrence (Simón et al., 2013; Te Beest et al., 2009). Using

¹The disease causing pathogen *Mycosphaerella graminicola* was recently named *Zymoseptoria tritici*. Others have used septoria tritici blotch (STB) to refer to septoria leaf blotch (SLB). Most published works from Ethiopia use the term SLB. Here, we use SLB to refer to the disease, and *M. graminicola* to refer to the disease causing pathogen.

environmental variables and different modeling approaches researchers have successfully predicted areas that are vulnerable to invasion from problematic species (Kumar et al., 2014, 2015; Macdonald et al., 2003; Schurich et al., 2014) including pathogens (Murray et al. 2011; Flory et al. 2012). More recently, Kumar (2014) predicted the occurrence of leaf rust in India using climatic variables, disease severity data, and multiple regression techniques. Despite the existence of highly reliable predictive techniques and the availability of global scale climatic, environmental, and topographic data, areas at risk of SLB disease have not been identified in Ethiopia. Likewise, the environmental conditions that favor high SLB disease severity levels have not been adequately evaluated. Here, we used disease severity data; linear regression and Boosted Regression Trees (BRT) modeling techniques; and suites of bioclimatic, remotely sensed, topographic, and soil variables to quantify the risk of SLB disease in central Ethiopia. Our objectives were to: a) map and predict SLB disease severity patterns in West Shewa, central Ethiopia; and b) identify the environmental variables associated with different levels of SLB disease severity. Our results can aid SLB disease monitoring and prevention efforts in Ethiopia including the identification and prioritization of vulnerable areas.

2. Materials and Methods

2.1 Study Area

West Shewa administrative zone is located in the central parts of Ethiopia within the Oromia regional state. It shares boundaries with Addis Ababa city, Amhara regional state, Southern Nations regional state, and four other Oromia administrative zones (i.e., East Shewa, East Welega, North Shewa, and Jimma; Fig. 1). Elevation in West Shewa ranges from 650m to 3,600m above sea level. The three dominant soil types in the study site are Nitosols , Vertisols,

and Cambisols (Batjes, 2012). Agricultural crops are mainly produced by subsistent farmers or by ‘small holding farmers’ for house hold consumption purposes. The top five cereal crops produced in the area include *teff*, barely, wheat, corn, and sorghum (CSA, 2011). In the study site, wheat is the third most important cereal crop, both in terms of area cultivated and quantity produced (Chamberlin and Schmidt, 2013; CSA, 2011). The projected Total Human Population (THP) of West Shewa in 2016 is about 1.4 million, while the projected THP of Ethiopia for the same period ranges from 101–104 million (CSA, 2013). Securing food for the growing THP in Ethiopia requires increasing wheat productivity and minimizing yield loss.

2.2 Disease Severity and Environmental Data

Geographically referenced SLB disease severity data were collected in 2014 from 64 different *kebeles* or ‘peasant associations’ (Takele et al., 2015b). The 64 *kebeles* were located in 11 *weredas* (districts) of the West Shewa administrative zone. The districts include Ambo, Ameya, Becho, Cheliya, Dawo, Dendi, Ejere, Kokir, Tikur Inchini, Weliso/ Goro, and Wenchi. The data were randomly collected at an interval of 5-10 km; thus, they fairly represent the environmental conditions of the study area (Fig. 1). Disease severity was recorded using the double digit standardized scoring system (Eyal, 1999; Saari and Prescott, 1975; Zhu et al., 2014). The first digit (D1) indicates the vertical advancement of the disease, while the second digit (D2) indicates disease symptoms in terms of damaged leaf area; the scale range for both D1 and D2 is 0-9. Percent disease severity was then calculated using the following equation:

$$\%Severity = (D1 / 9) * (D2 / 9) * 100$$

We selected a suite of remote sensing, bioclimatic, topographic, and soil variables to model SLB distribution in West Shewa (Appendix A). We considered MODIS normalized difference vegetation index (NDVI) for the *meher* growing season (i.e., June to December 2014)

to represent chlorophyll content, greenness and crop health. We chose the year 2014 because it represents our field data collection period (i.e., 2014). We considered monthly NDVI values, *meher* season average NDVI values, and their corresponding standard deviations. June, July, and August represent the environmental conditions of the wet season, which are crucial for the development of wheat crop and *M. graminicola*, while October and November represent the cool and dry season that favor SLB disease progress (Takele et al., 2015b).

Moisture- and temperature-related variables are known to influence the distribution of *M. graminicola* (Beyer et al., 2012; Pietravalle et al., 2003; Te Beest et al., 2009). Thus, we considered MODIS land surface temperature (LST), and the widely known 19 bioclimatic variables (Hijmans et al., 2005; <http://www.worldclim.org/>) for modeling *M. graminicola* in West Shewa (Appendix A). We normalized the MODIS LST products by subtracting LST values from their corresponding monthly maximum temperature values. We performed the normalization process to reduce elevation-induced variations from the LST products. Monthly maximum temperature values used in the normalization process were obtained from the WorldClim website (Hijmans et al., 2005).

All MODIS layers were obtained from National Aeronautics Space Agency (NASA) Land Processes Distributed Active Archive Center (LPDAAC) (Wan, 2013). We downloaded the data using the MODIS Reprojection Tool Web Interface (MRTWeb), and performed additional spatial analyses using ArcGIS software Version 10.3.1 (Esri, 2014). The MODIS NDVI layers had a spatial resolution of 232m X 232m, while the MODIS LST products had a spatial resolution of 1000m X 1000m. Elevation and Compound Topographic Index (CTI) were included to represent topographic heterogeneity. The topographic variables were derived from the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) at 90m spatial

resolution. Clay mass, water holding capacity, and total nitrogen content at the top 20cm horizon were considered as soil variables (Appendix A). We obtained all soil layers from the ISRIC-WISE database at 5 X 5 arc-minutes spatial resolution (Batjes, 2012). We performed spatial join in ArcGIS 10.3.1 (Esri, 2014), and linked the soil layers with the remote sensing and topoclimatic variables. A total of 50 predictor layers were prepared for modeling SLB disease severity in the study site (Appendix A). All spatial analyses were performed at the MODIS NDVI spatial resolution (i.e., 232m X 232m).

2.3 Disease Risk Modeling

We conducted multicollinearity tests among predictor variables in R statistical software using the ‘Hmisc’ package (R Development Core Team, 2011). We found most of the 50 variables to be highly correlated. We selected only 18 variables that were less correlated with each other (Pearson correlation coefficient $|r| < 0.80$) but were highly correlated with the response variable (i.e., disease severity). We assessed relationships among individual predictor variables and the response variable using linear regression models (Appendix B). Linear regression modeling was conducted in R using the ‘Mass’ and ‘Plotrix’ packages (R Development Core Team, 2011; Venables and Ripley, 2002). Concurrent to the linear modeling, we used all 18 predictor variables and ran a preliminary Boosted Regression Trees (BRT) model. Predictor layers that showed strong relationship with the response variable during linear regression also became influential in the BRT model.

Though simple and easy to interpret, linear regression models are not suited for detecting non-linear relationships. Thus, we selected the BRT modeling approach (implemented in the R “gbm” package) to conduct the final SLB disease modeling (Ridgeway, 2006). BRT models are

less sensitive to multicollinearity, can handle non-linear relationships, can fit non-parametric models, and thus have high predictive performance (De'ath and Fabricius, 2000; De'ath, 2007; Elith et al., 2008; Ridgeway, 2006; Sutton, 2004). We tested a number of settings and selected the following R settings for calibrating the final BRT model: number of trees = 100, learning rate = 0.001, tree complexity = 3, family = Gaussian, n.folds or cross-validation = 5, and bagging fraction = 0.5 (Appendices C, D). We selected a relatively small learning rate to allow the development of more than 1,000 trees (Elith et al., 2008). The disease severity data, collected in proportions, were transformed into logit scale; thus, assuming normal or Gaussian distribution was reasonable (Warton and Hut, 2011). The maximum number of predictor variables suggested for our data (64 disease severity records) is around six or seven. We used 10 influential variables in the final BRT model, and latter simplified the model to seven variables by dropping three less influential variables.

The seven variables used in the final and simplified BRT model were: June NDVI, November NDVI, mean growing season NDVI (all months considered), Tdifjune or Normalized LST for June (June maximum air temperature minus June land surface temperature), Tdifoct or Normalized LST for October (October maximum air temperature minus October land surface temperature), mean temperature of driest quarter (Bio9), and maximum temperature of warmest month (Bio5)² (Table 1). We fitted the best BRT model to the whole study site and assessed model fitness using different matrices such as deviance explained and standard error (SE). Predicted disease severity values were transformed back to proportions, exported to ArcGIS, and

²Several of our bioclimatic variables were highly correlated with each other, and less correlated with the response variable. The two exceptions are Bio9 and Bio5, which were correlated with the response variable, and with each other. We wanted to keep only one of these two predictors. Therefore, we ran different BRT models by leaving out one of them. Bio9 always became a top predictor, highly influencing the explanatory power of the BRT models. Though the influence of Bio5 was relatively low, it affected the explanatory power of the BRT models. When Bio5 was included, the R^2 values usually increased by 0.1. Elith et al. (2008) states that BRT can handle multicollinearity well. Thus, our decision to keep both Bio9 and Bio5 in the final BRT model is reasonable.

classified into five arbitrary disease severity classes. Additionally, model results were evaluated using coefficient of determination (R^2) by regressing the observed values (X-axis) on the predicted responses (Y-axis) (Warton and Hut, 2011). The presence of spatial autocorrelation in model residuals was tested using Moran's I (Fortin et al., 1989; Legendre and Fortin, 1989). The test was conducted in R using the 'ncf' package for every 10km distance increment (Bjørnstad, 2009).

3. Results

3.1 Effects of Environmental Variables on SLB Distribution

We found the MODIS NDVI, Normalized MODIS LST, and bioclimatic predictors to be highly influential in predicting SLB disease severity in West Shewa. June, mean, and November NDVI influenced the final BRT model by 20.1%, 16.1%, and 12.2%, respectively (Table 1 and Appendix E). Likewise, mean temperature of driest quarter (Bio9), October normalized LST, June normalized LST, and maximum temperature of warmest month (Bio5) influenced the final BRT model by 17.7%, 12.5%, 10.8%, and 10.6%, respectively (Table 1 and Appendix E). We found a high level of interaction among predictor variables. June NDVI interacted with November NDVI (interaction level = 2.46), while November NDVI interacted with mean temperature of driest quarter (Bio9) (interaction level = 0.76; Table 1). Predictors that had a relatively strong correlation with the response variable ($r \geq \pm 0.2$), consistently became the top predictors in both of the linear regression and BRT models (Figs. 2 and 3). The NDVI and bioclimatic variables had positive relationships, while the LST variables had negative relationships with SLB disease severity (Fig. 3). Topographic and soil predictors did not influence model results in this study.

3.2 Risk of SLB Disease in West Shewa

The BRT model identified and predicted high SLB disease severity in the central, northeastern and southeastern parts of the study site (Fig. 4). According to the model, the northwestern and southwestern parts of the study site are not severely affected (Fig. 4). Our final BRT map product highly agreed with the observed data, and with an independently created wheat map (10km X 10km grid that classifies wheat farms by different levels of productivity) (Fig. 4). We found linear relationships and a relatively high level of correlation among influential predictor variables and observed disease severity (Fig. 3). The Pearson's correlation (r) values were particularly higher for June NDVI (-0.26), mean NDVI (-0.28), and mean temperature of driest quarter (Bio9; -0.29) (Appendix B). The evaluation matrices generated by the final BRT model (simplified model) were sufficiently good; mean total deviance = 1.8, mean residual deviance = 0.76, SE = 0.1, number of trees = 7,500 (see Appendices C and D for all BRT outputs including evaluation matrices and R codes). The best BRT model had high predictive power; the R^2 value between observed and predicted disease severity was 0.63 with smaller confidence intervals ($P < 0.0001$; Fig. 5). No significant autocorrelation was detected in model residuals ($I \sim -0.05$, $P < 0.001$; Appendix F).

4. Discussion

The peak wheat production in West Shewa is estimated at 4,535 tons per grid (1 grid approximately equals 10km X 10km area) (You et al., 2014)³. The occurrence of SLB in this relatively productive region is a risk to wheat production and food security. As shown on the final map, high disease severity sites appear to be distinct and separate from low disease severity

³4,535 tons per grid = 45.35 quintal per hectare.

sites (Fig. 4). Our goal is to corroborate this particular observation (prediction) in the future by continuing the field data collection effort and conducting supplemental socio-economic surveys.

SLB disease severity levels in West Shewa were correlated with NDVI and LST values in June, October and November. The interaction detected among June and November NDVI values and mean temperature of driest quarter (Bio9; Table 1) indicates that the environmental/bioclimate conditions in June and November may strongly influence SLB disease distribution. Primary infection of *M. graminicola* usually occurs through deposition of airborne ascospores (O'Driscoll et al., 2014). With ascospores all around, the presence of host plants in June creates suitable conditions for SLB disease initiation. Moreover, cool temperature conditions in October and November create suitable conditions for secondary infection of SLB (Magboul et al., 1992; Takele et al., 2015b; Te Beest et al., 2009). Our findings suggest that monitoring crop health, moisture and land surface temperature conditions through MODIS NDVI, normalized LST and bioclimate variables may allow the accurate prediction of *meher* season SLB disease severity in Ethiopia. Here, we used only 1 year of disease severity data (the only data that were available to our team). We believe using multiple years of disease severity data and time-series analysis may produce better predictions by capturing the inter-annual variation in disease severity. MODIS NDVI and LST products are available at higher temporal resolution. Thus, time-series analysis can be performed in the future over a wider spatial extent (e.g., using weekly MODIS products for eastern and southern Africa).

The role of soil conditions on SLB disease distribution has not been clearly recognized in the literature (O'Driscoll et al., 2014; Simón et al., 2013). In our model, we considered soil type, water holding capacity, total nitrogen content, and clay mass in the top 20cm soil horizon (Batjes, 2012). Though the soil-related variables showed some level of correlation with the

response variable, they did not influence model results. Thus, all soil variables were not included in the final BRT model. We believe the coarse soil data available for the study site (10km X 10km) did not provide enough variation to discriminate *M. graminicola* infected wheat farms. We considered compound topographic index (CTI) and elevation to represent topographic variables. The CTI was not correlated with the response variable, and was dropped from the model early on. Elevation was correlated with the response variable; however, its relative influence was smaller than the remote sensing and bioclimatic predictors. Therefore, elevation was also excluded from the final BRT model. Elevation is known to influence the distribution of SLB particularly over a large study zone (O'Driscoll et al., 2014; Teferi and Gebreslassie, 2015; White et al., 2001). The relatively small size of our study site may have masked the role of elevation in SLB disease distribution. The role of soil, topographic heterogeneity, and climate variation on SLB disease distribution in Ethiopia requires further investigation. Relative humidity, wind direction and wind speed, which can influence SLB disease spread, may also be considered in future modeling efforts.

5. Conclusions

We demonstrated that BRT can be used to predict spatial patterns of SLB disease at different severity classes with reasonable accuracy. Compared to linear regression models, the BRT model produced better results, and detected interactions among predictor variables. Unlike soil and topographic related variables, remote sensing and bioclimatic variables highly influenced model results. Our findings suggest that June, October and November NDVI and LST values are highly important for mapping and predicting SLB disease in central Ethiopia. Using statistical and geospatial modeling techniques we produced SLB disease distribution maps at

different severity classes. We compared our map results with previously produced wheat yield maps and found high overlap. The results can be used for monitoring SLB disease and guiding future sampling efforts in central Ethiopia. Our methods can be adapted for quantifying disease severity and estimating yield loss for important crops and diseases. SLB disease poses a great threat to Ethiopia's wheat production. Ethiopia needs to manage fungal diseases to fulfil its food security ambitions. Our study demonstrates the integration of field and Earth observation data for monitoring septoria leaf blotch (*Septoria tritici*) disease severity in central Ethiopia.

Acknowledgments

We thank the Natural Resource Ecology Laboratory at Colorado State University for use of facilities. SK was supported by funding from the U.S. Geological Survey. Dr. Catherine Jarnevich, Dr. Sandy Cooper, and anonymous reviewers provided helpful comments that improved the manuscript. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government or Colorado State University.

References

- Abebe, T., Mehari, M., Legesse, M., 2015. Field response of wheat genotypes to *Septoria tritici* blotch in Tigray, Ethiopia. *J. Nat. Sci. Res.* 5, 146–153.
- Arraiano, L.S., Brown, J.K.M., 2006. Identification of isolate-specific and partial resistance to septoria tritici blotch in 238 European wheat cultivars and breeding lines. *Plant Pathol.* 55, 726–738. doi:10.1111/j.1365-3059.2006.01444.x.
- Batjes, N.H., 2012. ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2). ISRIC Rep. 56.
- Beyene, H., Franzel, S., Mwangi, W., 1990. Constraints to increasing wheat production, in: Tanner, D.G., Ginkel, M. Van, Mwangi, H. (Eds.), Mexico, pp. 284–293.
- Beyer, M., El Jarroudi, M., Junk, J., Pogoda, F., Dubos, T., Görgen, K., Hoffmann, L., 2012. Spring air temperature accounts for the bimodal temporal distribution of *Septoria tritici* epidemics in the winter wheat stands of Luxembourg. *Crop Prot.* 42, 250–255. doi:10.1016/j.cropro.2012.07.015.
- Broad, K., Agrawala, S., 2000. The Ethiopia food crisis- uses and ilmits of climate forecasts. *Science.* 289, 1693–1694.
- Bjørnstad, N., 2009. Package ncf: Spatial nonparametric covariance functions. Available at <<http://cran.r-project.org/web/packages/ncf/index.html>> (accessed June 1, 2016).
- Chamberlin, J., Schmidt, E., 2013. Ethiopian agriculture: A dynamic geographic perspective. *Food Agric. Ethiop. Prog. Policy Challenges* 21–54.
- CSA, 2015. Key findings of the 2014/2015 (2007 E.C.) agricultural sample surveys, CSA. Addis Ababa.
- CSA, 2013. Population projection of Ethiopia for all regions at wereda level from 2014 – 2017.

323 Addis Ababa.

324 CSA, 2011. Agricultural sample survey 2010/2011: Report on crop and livestock product
325 utilization season. Addis Ababa.

326 De'ath, G., 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88, 243–251.

327 De'ath, G., Fabricius, K., 2000. Classification and regression trees : A powerful yet simple
328 technique for ecological data analysis. *Ecology* 81, 3178–3192.

329 Dean, R., Van Kan, J., Pretorius, Z., Hammond-Kosack, K., Di Pietro, A., Spanu, P., Rudd, J.,
330 Dickman, M., Kahmann, R., Ellis, J., Foster, G., 2012. The top 10 fungal pathogens in
331 molecular plant pathology. *Mol. Plant Pathol.* 13, 414–430. doi:10.1111/j.1364-
332 3703.2011.00783.x.

333 Dercon, S., Hill, R.V., 2009. Growth from agriculture in Ethiopia: Identifying key constraints, 1–
334 70, available at http://users.ox.ac.uk/~econstd/Ethiopia%20paper%203_v5.pdf (accessed
335 June 1, 2016).

336 Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim.*
337 *Ecol.* 77, 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

338 ESRI, 2014. ArcGIS Desktop: Release 10.3. Redlands CA.

339 Flory, A. R., Kumar, S., Stohlgren, T. J., and Cryan, P. M. 2012. Environmental conditions
340 associated with bat White-Nose Syndrome mortality in the north-eastern United States. *J.*
341 *Appl. Ecol.* 49:680-689.

342 Fortin, M.-J., Drapeau, P., Legendre, P., 1989. Spatial autocorrelation and sampling design in
343 plant ecology. *Vegetatio* 83, 209–222. doi:10.1007/BF00031693.

344 Eyal, Z., 1999. The septoria tritici and stagonospora nodorum blotch diseases of wheat. *Eur. J.*
345 *Plant Pathol.* 105, 629–641. doi:10.1023/A:1008716812259.

346 Fones, H., Gurr, S., 2015. The impact of *Septoria tritici* Blotch disease on wheat: An EU
 347 perspective. Fungal Genet. Biol. 79, 3–7. doi:10.1016/j.fgb.2015.04.004.

348 Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution
 349 interpolated climate surfaces for global land areas. International Journal of Climatology
 350 25:1965-1978.

351 Kumar, S., L. G. Neven, and W. L. Yee. 2014. Assessing the potential for establishment of
 352 western cherry fruit fly using ecological niche modeling. J. Econ. Entomol. 107:1032-1044.

353 Kumar, S., Neven, L.G., Zhu, H., Zhang, R., 2015. Assessing the global risk of establishment of
 354 *Cydia pomonella* (Lepidoptera: Tortricidae) using CLIMEX and MaxEnt niche models. J.
 355 Econ. Entomol. 108, 1708–1719. doi:10.1093/jee/tov166.

356 Legendre, P., Fortin, M.J., 1989. Spatial pattern and ecological analysis. Vegetatio 80, 107–138.
 357 doi:10.1007/BF00048036.

358 Lyon, B., 2014. Seasonal drought in the Greater Horn of Africa and its recent increase during the
 359 March-May long rains. J. Clim. 27, 7953–7975. doi:10.1175/JCLI-D-13-00459.1.

360 Macauley, H., 2015. Cereal Crops: Rice, Maize, Millet, Sorghum, Wheat. Dakar, Senegal, Feed
 361 Africa, Background Paper, available at
 362 [http://www.afdb.org/fileadmin/uploads/afdb/Documents/Events/DakAgri2015/Cereal_Crop](http://www.afdb.org/fileadmin/uploads/afdb/Documents/Events/DakAgri2015/Cereal_Crops_-_Rice__Maize__Millet__Sorghum__Wheat.pdf)
 363 [s-_Rice__Maize__Millet__Sorghum__Wheat.pdf](http://www.afdb.org/fileadmin/uploads/afdb/Documents/Events/DakAgri2015/Cereal_Crops_-_Rice__Maize__Millet__Sorghum__Wheat.pdf) (accessed June 1, 2016).

364 Macdonald, I. a W., Reaser, J.K., Bright, C., Laurie, E., Howard, G.W., Murphy, S.J., Preston,
 365 G., 2003. Invasive alien species in Southern Africa: National Reports & Directory of
 366 Resources. The Global Invasive Species Programme. Cape Town, South Africa. 125p.

367 Magboul, A.M., Geng, S., Gilchrist, D.G., Jackson, L.F., 1992. Environmental influence on the
 368 infection of wheat by *Mycosphaerella Graminicola*. Ecol. Epidemiol. 82, 1407–1413.

369 Murray, K. A., Retallick, R. W. R., Puschendorf, R., Skerratt, L. F., Rosauer, D., McCallum, H.
 370 I., Berger, L., Speare, R., and VanDerWal, J. 2011. Assessing spatial patterns of disease risk
 371 to biodiversity: implications for the management of the amphibian pathogen,
 372 *Batrachochytrium dendrobatidis*. J. Appl. Ecol. 48:163-173.

373 O'Driscoll, A., Kildea, S., Doohan, F., Spink, J., Mullins, E., 2014. The wheat-Septoria conflict:
 374 A new front opening up? Trends Plant Sci. 19, 602–610. doi:10.1016/j.tplants.2014.04.011.

375 Pietravalle, S., Shaw, M.W., Parker, S.R., van den Bosch, F., 2003. Modeling of relationships
 376 between weather and *Septoria tritici* epidemics on winter wheat: A critical approach.
 377 Phytopathology 93, 1329–1339. doi:10.1094/PHYTO.2003.93.10.1329.

378 Ponomarenko, A., Goodwin, S.B., Kema, G.H.J., 2011. *Septoria tritici* blotch (STB) of wheat.
 379 Am. Phytopathol. Soc. 1–5. doi:DOI:10.1094/PHI-I-2011-0407-01.

380 R Development Core Team, R., 2011. R: A language and environment for statistical computing,
 381 R foundation for tsatistical computing. doi:10.1007/978-3-540-74686-7.

382 Ridgeway, G., 2006. gbm: Generalized boosted regression models. R package version 1(3).

383 Saari, E., Prescott, J., 1975. A scale for appraising the foliar intensity of wheat diseases. Plant
 384 Dis. Report. 59, 376–381.

385 Schurich, J., Kumar, S., Eisen, L., Moore, Chester, G., 2014. Modeling *Culex tarsalis* abundance
 386 on the northern Colorado front range using a landscape-level approach. J. Am. Mosq.
 387 Control Assoc. 20, 7–20. doi:10.2987/13-6373.1

388 Simón, M.R., Fleitas, M.C., Schalamuk, S., 2013. Recent advances on integrated foliar disease
 389 management with special emphasis in Argentina wheat rproduction. INTECH.
 390 doi:http://dx.doi.org/10.5772/51950.

391 Sutton, C.D., 2004. Classification and regression trees, bagging, and boosting. Handb. Stat. 24,

303–329. doi:10.1016/S0169-7161(04)24011-1.

Taffesse, A.S., Dorosh, P., Sinafikeh, A., Asrat, S., 2012. Crop rproduction in Ethiopia: Regional
Patterns and Trends, ESSP II Research Note. Washington, DC.

Takele, A., Lencho, A., Bekele, K., W/Ab, G., Hailu, E., 2015a. Estimated yield loss assessment
of bread wheat (*Triticumaestivum* L .) due to septoria leaf blotch *Septariatritici* (Roberge in
Desmaz) on Wheat in Holeta. Res. Plant Sci. 3, 61–67. doi:10.12691/plant-3-3-3.

Takele, A., Lencho, A., W/Ab, G., Hailu, E., Kassa, B., 2015b. Status of wheat Septoria Leaf
Blotch (*Septaria tritici* Roberge in Desmaz) in South West and Western Shewa Zones of
Oromiya Regional State , Ethiopia 3, 43–48. doi:10.12691/plant-3-3-1.

Tanner, D., Hulluka, M., 1991. Wheat research in Ethiopia: A historical pererspective.
IAR/CIMMYT, Addis Ababa, Ethiopia.

Te Beest, D.E., Shaw, M.W., Pietravalle, S., Van Den Bosch, F., 2009. A predictive model for
early-warning of Septoria leaf blotch on winter wheat. Eur. J. Plant Pathol. 124, 413–425.
doi:10.1007/s10658-009-9428-0.

Teferi, T.A., Gebreslassie, Z.S., 2015. Occurrence and intensity of wheat *Septoria tritici* blotch
and host response in Tigray, Ethiopia. Crop Prot. 68, 67–71.
doi:10.1016/j.cropro.2014.10.020.

Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S. Issues of Accuracy and
Scale 868. doi:10.1198/tech.2003.s33.

Kumar, P. V., 2014. Development of weather-based prediction models for leaf rust in wheat in
the Indo-Gangetic plains of India. Eur. J. Plant Pathol. 140, 429–440. doi:10.1007/s10658-
014-0478-6.

Wan, Z., 2013. MODIS Land Surface Temperature Products Users' Guide. Santa Barbara, CA.

415 Warton, D., Hut, F., 2011. The arcsine is asinine: the analysis of proportions in ecology. *Ecology*
 416 92, 3–10.

417 WFP, 2015. El Niño : Implications and scenarios for 2015, available at
 418 <http://documents.wfp.org/stellent/groups/public/documents/ena/wfp276236.pdf> (accessed
 419 June 1, 2016).

420 White, J.W., Tanner, D.G., Corbett, J.D., 2001. An agro-climatological characterization of bread
 421 wheat rproduction raeas in Ethiopia. *Inf. Syst.* 19.

422 You, L., Wood-Sichra, U., Fritz, S., Guo, Z., See, L., Koo, J., 2014. Spatial Production
 423 Allocation Model (SPAM) 2005 v2.0 [WWW Document]. URL
 424 <http://www.agrodep.org/dataset/mapspam-spatial-production-allocation-model-2005>
 425 (accessed June 1, 2016).

426 Zhu, Z., Bonnett, D., Ellis, M., Singh, P., Heslot, N., Dreisigacker, S., Gao, C., Mujeeb-Kazi, A.,
 427 2014. Mapping resistance to spot blotch in a CIMMYT synthetic-derived bread wheat. *Mol.*
 428 *Breed.* 1215–1228. doi:10.1007/s11032-014-0111-6.

429

Table 1. Interaction among variables and their relative influence in the final Boosted Regression Trees (BRT) model. November NDVI interacted with June NDVI and Bio9. June NDVI influenced the model by 20.1%. NDVI = Normalized Difference Vegetation Index; LST = Land Surface Temperature; Tmax = Maximum air temperature; Bio9 = mean temperature of driest quarter, and Bio5 = maximum temperature of warmest month.

Variable name	Relative Influence (%)	Interacted with	Interaction value
June NDVI	20.1	November NDVI	2.46
Bio9	17.7	November NDVI	0.76
Mean NDVI	16.1		
Tmax-LST June	12.5		
November NDVI	12.2		
Tmax-LST October	10.8		
Bio5	10.6

Figure 1. Map of the study site. The study area is located within the Oromia administrative region. Gam. = Gambela administrative region, B.G = Benishangul-Gumuz administrative region, S. Nations = Southern nations and nationalities administrative region.

Figure 2. Response curves of top six predictors based on the final Boosted Regression Trees (BRT) model. The response variable is shown on the y-axis. The relative influences of the variables are shown in percentages. Junendvi = NDVI values for June 2014, Bio9 = Mean temperature of driest quarter, Meanndvi = average NDVI values for 2014 (months considered were June to December), Tdifoct = normalized Land Surface Temperature (LST) of October 2014, Novndvi = NDVI for November 2014, Tdifjune = normalized LST of June 2014.

454

455

456 **Figure 3.** Relationships among top predictors and observed SLB disease severity (logit scale). a) = June NDVI, b) = mean NDVI, c) =
 457 mean temperature of driest quarter (Bio9), d) = Tmax - LST October, e) = Tmax - LST June, and f) = maximum temperature of
 458 warmest month (Bio5). NDVI = Normalized Difference Vegetation Index, Tmax = maximum air temperature of the month being
 459 considered, LST = Land Surface Temperature.

460

Figure 4. Septoria leaf blotch (SLB) disease severity map. Left: wheat growing areas in West Shewa classified by production levels. Data source: Spatial Production Allocation (SPAM) model. Right: SLB disease severity levels in West Shewa based on Boosted Regression Trees (BRT) model (this study). The dot density map showing black circles includes the field data used to develop the BRT model. According to the BRT model SLB disease-vulnerable areas exist in the central, northeastern and southeastern parts of the study site (indicated by warm colors).

Figure 5. Observed disease severity vs. predicted disease severity. We evaluated the final model by plotting the observed septoria leaf blotch (SLB) disease severity values on the predicted SLB values. We found a relatively high adjusted R^2 value (0.63) indicating that our prediction was fair and reliable. Minimum predicted value = 10.2, maximum predicted value = 90.1, mean predicted value = 46.2, and standard deviation = 20. The 90% confidence interval is shown by dotted lines; SI = severity index.

478 **Appendix A.** Variables used for modeling SLB disease, their spatial resolution and data sources.

No.	Variable name	Original Resolution	Data Source
1	Bio1 = Annual mean temperature	1km X 1km	WorldClim
2	Bio2 = Mean Diurnal Range	1km X 1km	WorldClim
3	Bio3 = Isothermality (BIO2/BIO7) (*100)	1km X 1km	WorldClim
4	Bio4 = Temperature Seasonality (standard deviation *100)	1km X 1km	WorldClim
5	Bio5 = Max Temperature of Warmest Month	1km X 1km	WorldClim
6	Bio6 = Min Temperature of Coldest Month	1km X 1km	WorldClim
7	Bio7 = Temperature Annual Range (BIO5-BIO6)	1km X 1km	WorldClim
8	Bio8 = Mean Temperature of Wettest Quarter	1km X 1km	WorldClim
9	Bio9 = Mean Temperature of Driest Quarter	1km X 1km	WorldClim
10	Bio10 = Mean Temperature of Warmest Quarter	1km X 1km	WorldClim
11	Bio11 = Mean Temperature of Coldest Quarter	1km X 1km	WorldClim
12	Bio12 = Annual Precipitation	1km X 1km	WorldClim
13	Bio13 = Precipitation of Wettest Month	1km X 1km	WorldClim
14	Bio14 = Precipitation of Driest Month	1km X 1km	WorldClim
15	Bio15 = Precipitation Seasonality (Coefficient of Variation)	1km X 1km	WorldClim
16	Bio16 = Precipitation of Wettest Quarter	1km X 1km	WorldClim
17	Bio17 = Precipitation of Driest Quarter	1km X 1km	WorldClim
18	Bio18 = Precipitation of Warmest Quarter	1km X 1km	WorldClim
19	Bio19 = Precipitation of Coldest Quarter	1km X 1km	WorldClim
20	Soil types	10km X 10km	ISRIC-WISE (FAO)
21	Water holding capacity (top 20cm soil horizon)	10km X 10km	ISRIC-WISE (FAO)
22	Clay mass (top 20cm soil horizon)	10km X 10km	ISRIC-WISE (FAO)
23	Total Nitrogen content (top 20cm soil horizon)	10km X 10km	ISRIC-WISE (FAO)
24	Elevation (m)	90m X 90m	SRTM DEM
25	Compound topographic index (CTI)	90m X 90m	SRTM DEM
26	June 2014 NDVI	232m X 232m	NASA LPDAAC
27	July 2014 NDVI	232m X 232m	NASA LPDAAC
28	August 2014 NDVI	232m X 232m	NASA LPDAAC
29	September 2014 NDVI	232m X 232m	NASA LPDAAC
30	October 2014 NDVI	232m X 232m	NASA LPDAAC
31	November 2014 NDVI	232m X 232m	NASA LPDAAC
32	December 2014 NDVI	232m X 232m	NASA LPDAAC
33	Average (July - December 2014) NDVI	232m X 232m	NASA LPDAAC
34	Standard Deviation (July - December 2014) NDVI	232m X 232m	NASA LPDAAC
35	June 2014 LST	1km X 1km	NASA LPDAAC
36	July 2014 LST	1km X 1km	NASA LPDAAC
37	August 2014 LST	1km X 1km	NASA LPDAAC
38	September 2014 LST	1km X 1km	NASA LPDAAC
39	October 2014 LST	1km X 1km	NASA LPDAAC
40	November 2014 LST	1km X 1km	NASA LPDAAC

41	December 2014 LST	1km X 1km	NASA LPDAAC
42	Average (July - December 2014) LST	1km X 1km	NASA LPDAAC
43	Standard Deviation (July-December 2014) LST	1km X 1km	NASA LPDAAC
44	Tdifjune (Tmax June - June 2014 LST)	1km X 1km	NASA /WorldClim
45	Tdifjuly (Tmax July - July 2014 LST)	1km X 1km	NASA /WorldClim
46	Tdifaug (Tmax August - August 2014 LST)	1km X 1km	NASA /WorldClim
47	Tdifsept (Tmax September - September 2014 LST)	1km X 1km	NASA /WorldClim
48	Tdifoct (Tmax October - October 2014 LST)	1km X 1km	NASA /WorldClim
49	Tdinov = (Tmax November - November 2014 LST)	1km X 1km	NASA /WorldClim
50	Tdifdec (Tmax December - December 2014 LST)	1km X 1km	NASA /WorldClim

479

480 MODIS = Moderate Resolution Imaging Spectroradiometer; NDVI = Normalized Difference Vegetation

481 Index; LST = Land Surface Temperature; Tmax = Maximum air temperature; NASA = National Aeronautics

482 and Space Administration, LPDAAC = Land Processes Distributed Active Archive Center, SRTM =

483 Shuttle Radar Topography Mission, DEM = Digital Elevation Model. All MODIS, LST, and SRTM

484 products were obtained from the LPDAAC webpage: <https://lpdaac.usgs.gov/>.

485

Appendix B. Pearson correlation (*r*) test results of the seven variables used in the final BRT model (*n* = 64)

	SI	June.ndvi	Nov.ndvi	Mean.ndvi	Tdif.june	Tdif.oct	Bio5	Bio9
SI	1							
June.ndvi	-0.26	1						
Nov.ndvi	-0.2 ^{ns}	0.58	1					
Mean.ndvi	-0.28	0.56	0.38	1				
Tdif.june	0.21 ^{ns}	-0.66	-0.67	-0.18 ^{ns}	1			
Tdif.oct	0.26	-0.58	-0.59	-0.14 ^{ns}	0.8	1		
Bio5	-0.24 ^{ns}	-0.21 ^{ns}	-0.54	0.09 ^{ns}	0.45	0.48	1	
Bio9	-0.29	-0.14 ^{ns}	-0.43	0.10 ^{ns}	0.34	0.38	0.95	1

SI = Severity Index in logit scale; *ns* indicates non-significance at alpha = 0.05. All other correlation results were significant at alpha = 0.05.

Appendix C. Number of trees. The simplified model had 7300 trees. Our response variable was named Tsimen which stands for the mean severity index in logit scale (transformed mean severity index); tree complexity (d) = 3, and learning rate lr = 0.001.

```

497 Appendix D. Detailed Boosted Regression Trees (BRT) model results and R codes.
498
499 # Library 'dismo', 'gbm', and 'plotrix' are needed. The gbm package will import other needed
500 packages such as 'raster'. Codes are adapted for Elith et al. (2008).
501 # Response is in column 7, data are from column 8 to 17
502 # The training file is named Training3, response is in logit scale, training data is in .csv format
503 # Ran BRT model
504 > brtmodel.1<- gbm.step(data=Training3,gbm.x = 8:17,gbm.y = 7, family =
505 "gaussian",tree.complexity = 3,n.folds = 5, n.trees=100,learning.rate = 0.001,bag.fraction = 0.5)
506 # Results
507 brtmodel.1<- gbm.step(data=Training3,gbm.x = 8:17,gbm.y = 7, family =
508 "gaussian",tree.complexity = 3,n.folds = 5, n.trees=100,learning.rate = 0.001,bag.fraction = 0.5)
509 GBM STEP - version 2.9
510 Performing cross-validation optimization of a boosted regression tree model
511 for Tsimean and using a family of gaussian
512 Using 64 observations and 10 predictors
513 creating 5 initial models of 100 trees
514 folds are unstratified
515 total mean deviance = 1.8806
516 tolerance is fixed at 0.0019
517 ntrees resid. dev.
518 100 1.9116
519 now adding trees...
520 200 1.8825
521 300 1.8552
522 400 1.8308
523 500 1.8076
524 600 1.7907
525 700 1.7741

```

526	800	1.7597
527	900	1.7437
528	1000	1.7302
529	1100	1.7204
530	1200	1.7119
531	1300	1.701
532	1400	1.6922
533	1500	1.685
534	1600	1.6801
535	1700	1.6735
536	1800	1.6694
537	1900	1.6668
538	2000	1.664
539	2100	1.6625
540	2200	1.6603
541	2300	1.6592
542	2400	1.6584
543	2500	1.6544
544	2600	1.6557
545	2700	1.6538
546	2800	1.6528
547	2900	1.6524
548	3000	1.6512
549	3100	1.6486
550	3200	1.6509
551	3300	1.6512
552	3400	1.6501
553	3500	1.6512


```

554 3600 1.6517
555 3700 1.654
556 3800 1.6538
557 3900 1.6553
558 4000 1.6558
559 4100 1.6575
560 fitting final gbm model with a fixed number of 3100 trees for Tsimean
561 mean total deviance = 1.881
562 mean residual deviance = 0.97
563 estimated cv deviance = 1.649; se = 0.101
564 training data correlation = 0.729
565 cv correlation = 0.39; se = 0.079
566 elapsed time - 0.06 minutes
567 # Summarize model and see results
568 > summary(brtmodel.1)
569 Bio9      Bio9 20.399019
570 Junendvi Junendvi 16.538307
571 Bio5      Bio5 11.462789
572 Meanndvi Meanndvi 11.270890
573 Tdifoct   Tdifoct 10.557773
574 Novndvi   Novndvi 9.702041
575 Tdifjune  Tdifjune 6.827672
576 Tdifnov   Tdifnov 5.067511
577 Augndvi   Augndvi 5.017154
578 Octndvi   Octndvi 3.156844
579 # Then simplify model by dropping three less contributing variables
580 > brtsimp1 <- gbm.simplify(brtmodel.1, n.drops = 3)
581 > brtsimp2<- gbm.step(data=Training3,gbm.x = brtsimp1$pred.list[[3]],gbm.y = 7,family =
582 "gaussian", tree.complexity = 3,n.folds = 5,n.trees=100,learning.rate = 0.001,bag.fraction = 0.5)

```

```

583 # Fitting simplified model to data
584 brtsimp2<- gbm.step(data=Training3,gbm.x = brtsimp1$pred.list[[3]],gbm.y = 7,family =
585 "gaussian", tree.complexity = 3,n.folds = 5,n.trees=100,learning.rate = 0.001,bag.fraction = 0.5)
586 GBM STEP - version 2.9
587 Performing cross-validation optimization of a boosted regression tree model
588 for Tsimean and using a family of gaussian
589 Using 64 observations and 7 predictors
590 creating 5 initial models of 100 trees
591 folds are unstratified
592 total mean deviance = 1.8806
593 tolerance is fixed at 0.0019
594 ntrees resid. dev.
595 100 1.9973
596 now adding trees...
597 200 1.9575
598 300 1.9274
599 400 1.898
600 500 1.8717
601 600 1.8424
602 700 1.8157
603 800 1.7952
604 900 1.7774
605 1000 1.7612
606 1100 1.7472
607 1200 1.734
608 1300 1.7159
609 1400 1.7068
610 1500 1.7025
611 1600 1.6952

```

612	1700	1.6881
613	1800	1.6824
614	1900	1.6772
615	2000	1.6692
616	2100	1.6622
617	2200	1.6585
618	2300	1.6557
619	2400	1.6502
620	2500	1.6486
621	2600	1.6464
622	2700	1.6426
623	2800	1.6422
624	2900	1.6372
625	3000	1.6346
626	3100	1.6325
627	3200	1.6322
628	3300	1.6292
629	3400	1.628
630	3500	1.6261
631	3600	1.625
632	3700	1.6222
633	3800	1.6223
634	3900	1.6228
635	4000	1.6186
636	4100	1.6182
637	4200	1.6159
638	4300	1.6153
639	4400	1.6127

640	4500	1.6122
641	4600	1.6108
642	4700	1.6091
643	4800	1.6075
644	4900	1.6075
645	5000	1.6063
646	5100	1.6075
647	5200	1.607
648	5300	1.6062
649	5400	1.6074
650	5500	1.606
651	5600	1.6057
652	5700	1.6067
653	5800	1.6037
654	5900	1.6044
655	6000	1.604
656	6100	1.6035
657	6200	1.6043
658	6300	1.6016
659	6400	1.6002
660	6500	1.6007
661	6600	1.6017
662	6700	1.6021
663	6800	1.6016
664	6900	1.6017
665	7000	1.6023
666	7100	1.603
667	7200	1.6025

```

668 7300 1.6002
669 7400 1.6014
670 7500 1.602
671 fitting final gbm model with a fixed number of 7300 trees for Tsimean
672 mean total deviance = 1.881
673 mean residual deviance = 0.762
674 estimated cv deviance = 1.6; se = 0.215
675 training data correlation = 0.785
676 cv correlation = 0.43; se = 0.118
677 elapsed time - 0.13 minutes
678 # Summarize simplified model and see results
679 > summary(brtsimp2)
680 Junendvi.....Junendvi 20.08604
681 Bio9      Bio9 17.74339
682 Meanndvi Meanndvi 16.08784
683 Tdifoct   Tdifoct 12.47841
684 Novndvi   Novndvi 12.16231
685 Tdifjune  Tdifjune 10.80997
686 Bio5      Bio5 10.63204
687 > save.image("C:/R/BrWork_3.RData")
688 # Looking for interactions using the simplified model
689 > find.int <- gbm.interactions(brtsimp2)
690 gbm.interactions - version 2.9
691 Cross tabulating interactions for gbm model with 7 predictors
692 1 2 3 4 5 6
693 > find.int$interactions
694 Junendvi Novndvi Meanndvi Tdifjune Tdifoct Bio5 Bio9
695 Junendvi    0  2.46   0.01   0.02   0.07 0.04 0.14

```

```

696 Novndvi      0  0.00  0.48  0.01  0.01 0.32 0.76
697 Meanndvi     0  0.00  0.00  0.02  0.00 0.22 0.24
698 Tdifjune     0  0.00  0.00  0.00  0.06 0.09 0.17
699 Tdifoct      0  0.00  0.00  0.00  0.00 0.05 0.15
700 Bio5         0  0.00  0.00  0.00  0.00 0.00 0.00
701 Bio9         0  0.00  0.00  0.00  0.00 0.00 0.00
702 # Rank interactions
703 > find.int$rank.list
704 var1.index var1.names var2.index var2.names int.size
705 1      2  Novndvi      1  Junendvi  2.46
706 2      7   Bio9       2  Novndvi   0.76
707 # Plotting and saving plots to folder (based on simplified model)
708 gbm.plot(brtsimp2,n.plots=7,write.title = FALSE,plot.layout=c(2,4))# plots all 7
709 gbm.plot(brtsimp2,n.plots=6,write.title = FALSE,plot.layout=c(2,3))# plots only 6 of them, first
710 remove the plots!!!
711 # Forecasting/predicting (based on simplified model)
712 # First import the data to be predicted, the file is named Predicting3 and the column names are
713 same as Training3,
714 # The output is named forecast3
715 forecast3 <- predict.gbm(brtsimp2, Predicting3,n.trees=brtsimp2$gbm.call$best.trees,
716 type="response")
717 # to back transform forecasted values to percentage values use
718 forecasttrans3<-(inv.logit(forecast3))*100 # first load library boot, it is a system package.
719 # Writing or saving the back transformed prediction as final forecast named 'finalforecast3.csv'
720 write.csv(forecasttrans3,file ='finalforecast3.csv')
721 # Predicted severity index is saved as 'finalforecast3.csv' in the current directory
722 # Export the final predicted ".csv" file to ArcGIS for further processing
723

```

Appendix E. Relative influence. The relative influence of the seven variables used in the final Boosted Regression Trees (BRT) model are shown. NDVI = Normalized Difference Vegetation Index, LST = Land Surface Temperature, Tmax = maximum air temperature (monthly), Bio9 = mean temperature of driest quarter and Bio5 = maximum temperature of warmest month.

Appendix F. Results of test for spatial autocorrelation in model residuals.

Appendix F-1. Correlogram shows the Moran's I values for every 10km (10,000m) distance.

The average Moran's I is ~ -0.05 which is very close to zero. Open circles in the graph indicate

no significant autocorrelation in model residuals.

Appendix F-2. Table shows detailed Moran's I results and their associated P values.

No	Distance (Km)	Morans' I	P
1	10	-0.09	0.19
2	20	0.00	0.39
3	30	0.00	0.42
4	40	-0.03	0.38
5	50	-0.01	0.49
6	60	-0.03	0.39
7	70	0.02	0.30
8	80	-0.02	0.49
9	90	0.07	0.19
10	100	-0.13	0.16
11	110	0.07	0.25
12	120	0.09	0.32
13	130	-0.55	0.13