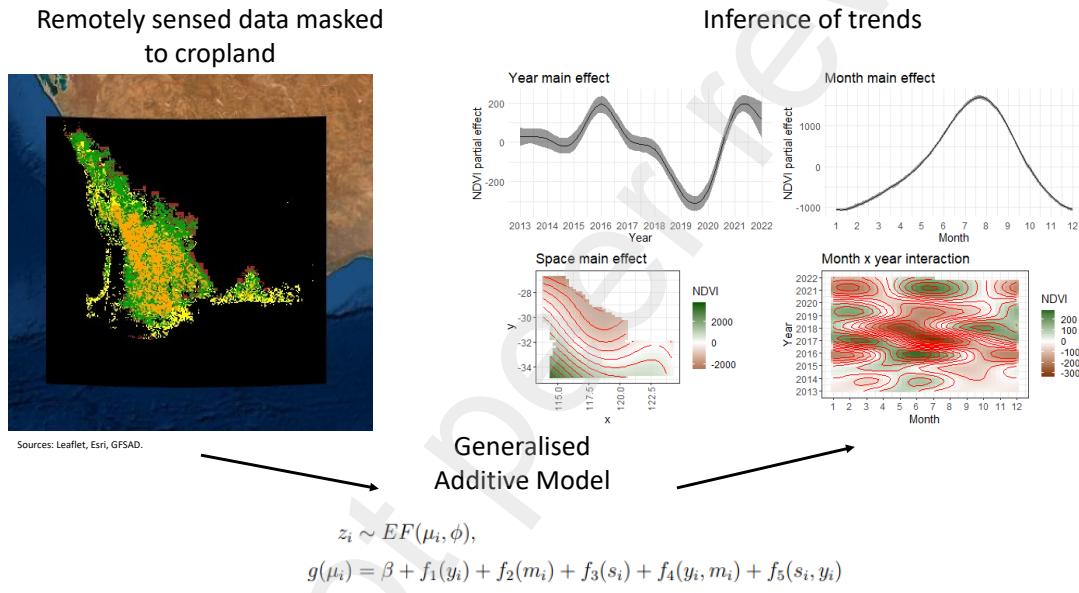


¹ Graphical Abstract

² A framework for modelling spatio-temporal trends in crop production using generalised additive models

⁴ Michael J Wellington, Roger Lawes, Petra Kuhnert



⁵ Highlights

⁶ **A framework for modelling spatio-temporal trends in crop produc-**
⁷ **tion using generalised additive models**

⁸ Michael J Wellington, Roger Lawes, Petra Kuhnert

⁹ • A modelling framework is proposed for inferring trends in crop produc-

¹⁰ tion.

¹¹ • The model accounts for inter-annual, spatial, and seasonal variation.

¹² • Predictions correlated with production data from Western Australia
¹³ ($p=0.03$).

¹⁴ • Cropland NDVI decreased by 2.5% per year from 2014 to 2021 in Mada-
¹⁵ gascar.

16 A framework for modelling spatio-temporal trends in
17 crop production using generalised additive models

18 Michael J Wellington^a, Roger Lawes^b, Petra Kuhnert^c

^a*Fenner School of Environment and Society, Australian National University, B141, B48,
B48A, Linnaeus Way, Canberra, 2601, ACT, Australia*

^b*CSIRO Agriculture & Food, Underwood Ave, Floreat, 6014, WA, Australia*

^c*CSIRO Data61, 41 Boggo Road, Dutton Park, 4102, QLD, Australia*

19 **Abstract**

20 Satellite imagery provides opportunities for inference of trends in crop pro-
21 duction across space and time. However, the large size of these datasets has
22 made application of statistical modelling approaches computationally diffi-
23 cult. Recent advances in computational techniques and infrastructure have
24 allowed generalised additive models to be fitted to very large datasets. We
25 propose a framework for inferring trends in crop production across space
26 and time using generalised additive models which accounts for inter-annual
27 trends (main effect of year), spatial distribution (main effect of space), crop
28 ontogeny (main effect of month), inter-annual changes in seasonality (inter-
29 action between year and month), and inter-annual changes in spatial distri-
30 bution (interaction between year and space). Application of the proposed
31 model to farm scale, multi-field sites in the Ord River Irrigation Area, West-
32 ern Australia, demonstrates that this approach is able to decompose variation
33 into the aforementioned effects. Furthermore, comparison of grain produc-
34 tion observations and estimates for the Western Australian Wheatbelt as
35 ground-truth data showed agreement with inferences drawn from the pro-
36 posed model, with prediction terms for the main effect of year positively
37 correlated with estimated tonnes produced from 2013 to 2021 ($p = 0.03$).
38 Finally, application to Madagascar, which has been experiencing a food cri-

39 sis, revealed a decreasing trend in cropland Normalised Difference Vegetation
40 Index (NDVI) from 2014 to 2021 of 2.5%, raising concerns about ongoing
41 food security. The proposed modelling framework is adaptable to numerous
42 agricultural research problems.

43 *Keywords:* remote sensing, generalised additive models, agricultural
44 productivity, spatio-temporal

45 **1. Introduction**

46 Estimating trends in crop production over both space and time informs
47 responses to important food security and environmental challenges (Pretty,
48 1997; Reardon et al., 1995; Rockström et al., 2017). The ubiquity of remote
49 sensing data and low cost of acquisition means it can be easily applied to
50 cropland monitoring over space and time (CGIAR, 2019). Much of the focus
51 in recent big data initiatives has been on machine learning approaches for
52 prediction of outcomes such as crop yield (Holloway and Mengersen, 2018;
53 Lobell et al., 2015; Ouattara et al., 2020). However, approaches that enable
54 inference of spatio-temporal patterns remain important to develop for big
55 data and the remote sensing community. We sought to investigate the capa-
56 bility of generalised additive models (GAMs), as a semi-parametric statisti-
57 cal approach, to decompose spatio-temporal data into inter-annual, seasonal,
58 and spatial changes at various scales, enabling inference of trends.

59 Institutions and service providers are often interested in how crop produc-
60 tion has trended over time, as this can demonstrate yield progress or regress
61 and inform research and development investment (Cassman and Grassini,
62 2013; Fischer et al., 2014; Dossou-Yovo et al., 2020). However, overall changes
63 in productivity may correspond with changes in the seasonal or spatial dis-
64 tribution of productivity. For example, there has been historical interest in
65 how trends towards earlier sowing dates in the Western Australian Wheatbelt
66 relate to crop yields, representing a change in seasonal distribution of crop
67 production (Stephens and Lyons, 1998; Fletcher et al., 2016). Conversely, the

68 current southern Madagascar food crisis has been attributed to the greater
69 biophysical propensity of that region to experience drought relative to the
70 north, in addition to socioeconomic factors, representing spatial variation be-
71 tween years (Rakotoarison et al., 2021). It is therefore important to quantify
72 the temporal and spatial components within crop production trends so that
73 changes in agricultural systems can be comprehensively understood.

74 Software applications for remote sensing applications also support this
75 dichotomy between simple, pixel-based, descriptive approaches and complex
76 predictive approaches. For example, Google Earth Engine provides a use-
77 ful access and computation platform for geospatial data. Several functions
78 are available for reducing image collections over time and identifying trends
79 including linear regression, linear fit, and ridge regression (Gorelick et al.,
80 2017). Conversely, more complex classification algorithms such as support
81 vector machines and random forest are available as Earth Engine functions,
82 and incorporation of TensorFlow methods is also possible (Abadi et al., 2016).
83 Additive decomposition of time-series for detection of trends and breakpoints
84 in NDVI has been developed for R software packages (Forkel et al., 2013;
85 Verbesselt et al., 2010). However, these are applied to either time-series of
86 pixels or spatially aggregated time-series meaning it is difficult to infer con-
87 tributions of both space and time holistically, and therefore derive measures
88 of variance and uncertainty across these dimensions. This reveals a gap in
89 approaches for statistical inference of spatio-temporal trends from remote
90 sensing data, such as generalised linear, generalised additive, and associated
91 mixed model approaches.

92 Wikle et al. (2019) provide a comprehensive overview of methods for
93 spatio-temporal data. They range from descriptive methods that allow spa-
94 tial and temporal variation to be explored, exploratory methods that in-
95 clude GAMs, covariance based methods including kriging which are useful
96 for prediction and more complex statistical models that adopt a Bayesian
97 Hierarchical Modelling framework. For large data, GAMs have proven to

98 be a challenging modelling paradigm due to computational considerations
99 in structuring the smoothers to capture the spatio-temporal dynamics of the
100 data. Further, the ability of a GAM to capture complex dynamical properties
101 of a system to aid inference is still under investigation.

102 Despite not being widely applied to remote sensing data, GAMs are a
103 popular method for spatio-temporal analyses of point-based environmental
104 and agricultural data. They have been a common method of spatio-temporal
105 analysis in forestry, fisheries, and ecology (Fewster et al., 2000; Venables
106 and Dichmont, 2004; Simpson, 2018; Yee and Mitchell, 1991). Furthermore,
107 GAMs have been applied to wheat yield prediction based on in-situ yield data
108 combined with weather and geolocation information in Western Australia
109 (Chen et al., 2019). The large size of remote sensing datasets has limited the
110 application of GAMs. However, recent advances in the statistical computa-
111 tion of GAMs for very large datasets have enabled application to datasets
112 derived from remote or proximal sensing imagery (Li and Wood, 2020; Wood
113 et al., 2015, 2017). De Rosa et al. (2021) found that GAMs outperformed
114 random forest in reproducing spatial variability of pasture biomass estimates
115 from unmanned aerial vehicle imagery, demonstrating the capability of this
116 approach for prediction and inference of spatial variation.

117 In this paper, we aim to demonstrate that a GAM framework can en-
118 able inference of crop production trends and variation in space and time at
119 various scales. To this end, we applied the proposed framework to several
120 multi-field scale analyses representing seasonal and spatial variation in time.
121 We then applied the framework at regional scale to the Western Australian
122 Wheatbelt, a well-studied grain production region, and compared model out-
123 puts with real production figures and observations. Finally, we applied the
124 framework to the nation of Madagascar, which has been experiencing a se-
125 vere famine since mid-2021, with the World Food Programme (2022) report-
126 ing that 309,000 children are projected to suffer from acute malnutrition
127 through to August 2022. We use the model outputs to draw inferences on

₁₂₈ crop production patterns that led to the food crisis.

₁₂₉ **2. Methods**

₁₃₀ *2.1. Proposed model*

₁₃₁ We propose a GAM with a nested main effect and interaction structure
₁₃₂ for space and time:

$$\begin{aligned} z_i &\sim EF(\mu_i, \phi), \\ g(\mu_i) &= \beta + f_1(y_i) + f_2(m_i) + f_3(s_i) + f_4(y_i, m_i) + f_5(s_i, y_i) \end{aligned} \tag{1}$$

₁₃₃ where z_i is a response associated with crop production at the i th pixel,
₁₃₄ such as NDVI, $EF(\mu_i, e)$ represents an exponential family distribution with
₁₃₅ mean, μ_i and scale parameter, ϕ , g is the link function, β is an intercept,
₁₃₆ and $f_j(j = 1, \dots, 5)$ are smooth functions that incorporate combinations of
₁₃₇ terms m_i , y_i and s_i that represent month, year and spatial terms respectively.
₁₃₈ The model was fitted with a Gaussian error term following an autoregressive
₁₃₉ (AR) process at each site, A description for each effect and a guide to their
₁₄₀ interpretation when applied to an agricultural system is given in Table 1.

₁₄₁ GAMs use specified functions for smoothing splines with flexibility deter-
₁₄₂ mined by the basis dimension k , which effectively acts as the upper limit on
₁₄₃ degrees of freedom for each smoothing spline (Wood, 2006). The smoothing
₁₄₄ splines and basis dimensions set to each term are shown in Table 2.

Term	Statistical Interpretation	Agricultural System Interpretation
Main effect of year: $f_1(y_i)$	The effect of year across space and months.	This effect enables inference of overall inter-annual trends in agricultural system productivity.
Main effect of month: $f_2(m_i)$	The effect of month across space and years.	Crop ontogeny for the period modelled.
Main effect of space: $f_3(s_i)$	The effect of space, across months and years.	Inherent spatial variation in crop production.
Month \times Year interaction: $f_4(y_i, m_i)$	The effect of month between years and across space. This smooth can be interpreted as the alterations to be applied to $f_2(m_i)$ as year varies, or vice versa.	This term enables inference of changes in crop seasonality over time.
Space \times Year interaction: $f_5(s_i, y_i)$	The effect of space between years and across months. This smooth can be interpreted as the alterations to be applied to $f_3(s_i)$ as year varies, or vice versa.	This term enables inference of changes in the spatial distribution of crop productivity over time.

Table 1: Terms included in the proposed GAM framework and their interpretation when applied to an agricultural system.

Term	Smoothing spline	Basis dimension (k)
$f_1(y_i)$	Cubic regression (Royston and Sauerbrei, 2007)	n unique years
$f_2(m_i)$	Cyclic cubic regression (Royston and Sauerbrei, 2007; Simpson, 2018)	n unique months
$f_3(s_i)$	Gaussian process (Paciorek, 2003; Wood, 2006)	50
$f_4(y_i, m_i)$	Cubic regression (Royston and Sauerbrei, 2007), cyclic cubic regression (Royston and Sauerbrei, 2007; Simpson, 2018)	n unique years, n unique months
$f_5(s_i, y_i)$	Gaussian process (Paciorek, 2003; Wood, 2006), cubic regression (Royston and Sauerbrei, 2007)	50, n unique years

Table 2: Terms included in the proposed GAM framework, the smoothing splines, and basis dimensions k used in model fit.

145 *2.2. Model fit and interpretation*

146 The **bam** function from the **mgcv** R package was used to fit GAMs with
 147 discretised covariates in R version 4.1.3 (Hastie and Tibshirani, 2017; Li and
 148 Wood, 2020; R Core Team, 2022; Wood et al., 2015, 2017; Wood and Wood,
 149 2015). The model for the Madagascar dataset was run on a virtual instance of
 150 R 4.1.3 with 16 virtual central processing units (vCPUs) and 128 Gigabytes
 151 of random access memory (RAM) (Edmondson, 2019). All models were first
 152 fitted assuming independent residuals and an autocorrelation function was
 153 calculated to determine the ρ value to use in the subsequently fitted AR1
 154 model. This approach followed the process and functions in the **itsadug**
 155 R package for correcting temporal autocorrelation in GAMs fitted to time-

156 series (Van Rij et al., 2015). A linear trend was estimated for the Madagascar
157 dataset from 2014 to 2021 by fitting year as a linear term to a log-transformed
158 NDVI response, with all other smooths remaining the same.

159 Spatio-temporal trends were visualised with a partial effects plot of each
160 smooth in Eq. 1, using data from the `mgcv` package passed to `ggplot2`
161 (Fasiolo et al., 2020a,b; Wickham et al., 2019). Model fit was evaluated using
162 the check function in `mgcv` and `mgcv` alongside semi-variogram plots of the
163 residuals to check for spatial dependence and plots of the autocorrelation
164 function to examine any temporal dependencies (Van Rij et al., 2015).

165 The loss in deviance explained attributable to specific terms for the Ord
166 River scenarios were calculated by iterating through model fits with selected
167 terms removed. All terms containing ‘year’; that is the year main effect and
168 its interactions with both month and space, were removed to estimate the loss
169 in deviance attributable to the overall effect of year. This is because it is not
170 possible to exclude a main effect without an interaction effect compensating
171 for its loss. The loss in deviance explained attributable to the interaction
172 terms (Table 1) were estimated by dropping those terms individually.

173 Inferences were drawn from the Western Australian Wheatbelt model
174 from plotting the partial effects of each smooth in Eq. 1, in addition to
175 inspecting model predictions for years of interest. Further, grain production
176 estimates made at end of season for Western Australia were compared with
177 model terms for the main effect of year using a Spearman correlation test.

178 *2.3. Study areas and data collection*

179 *2.3.1. Multi-field scale sites*

180 Three scenarios which represented variation between years, years and
181 months, and in space between years were selected from visual inspection of
182 a time-series of images of the Ord River Irrigation Area, Western Australia.
183 Growth of a perennial tree crop was chosen to demonstrate variation between
184 years, a shift from perennial to annual cropping was chosen for variation in

185 year and month, and expansion of irrigated land was chosen to represent
186 a change in space between years. These scenarios are illustrated in Figure
187 1 with true-colour images from selected points across the time-series. The
188 perennial growth scenario shows increasing greenness, the perennial to an-
189 nual scenario shows a perennial crop, then cultivation prior to growth of an
190 annual crop, and the change in space (spatial dynamics) scenario shows na-
191 tive vegetation before the western area of the site is developed for irrigated
192 cropping, and the eastern portion follows.

193 Analysis ready Landsat-8 surface reflectance data were acquired from
194 the Digital Earth Australia platform for each site for the date range 1st
195 January 2013 to 31st December 2021 (Krause et al., 2021; Dhu et al., 2017).
196 This period gave 197 possible timesteps from Landsat-8. The number of
197 timesteps remaining in the final datasets after the minimum good-quality
198 pixel threshold was set to 99% in the function for loading analysis ready
199 data are shown in Table 3. NDVI was calculated as:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad (2)$$

200 where NIR is the near infrared, band 5, of Landsat-8 and red is band 4,
201 using the index calculation function in Digital Earth Australia (Krause et al.,
202 2021; Rouse et al., 1973). This meant the NDVI ranged from 0 to 1. Table
203 3 summarises each site and dataset.

Scenario	n time (out of a possible 197 time-steps)	n pixel	n total
Perennial growth	133	5,928	788,424
Perennial to annual	144	1,028	148,032
Spatial dynamics	134	7,928	1,062,316

Table 3: Data size of NDVI collection for proof of concept scenario sites.

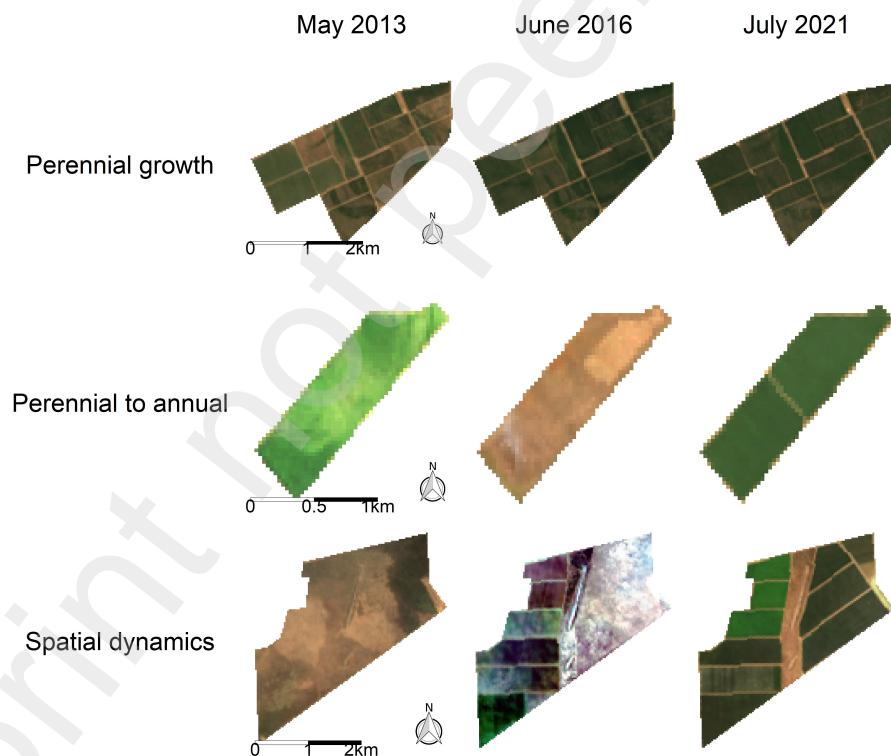


Figure 1: True colour images of each site at the Ord River Irrigation Area used for proof of concept spatio-temporal modelling. Column labels show date of image acquisition while row labels show the biophysical effect of each site was chosen to demonstrate.

204 *2.3.2. Regional and national scale sites*

205 The well-studied and described Western Australian Wheatbelt was used
206 as a preliminary study site for comparison with on-ground grain produc-
207 tion estimates and observations published by the Grain Industry Association
208 of Western Australia. These estimates are also informed by remote sens-
209 ing assessments in addition to cross-checking with agronomists, agribusiness
210 consultants, the Western Australian Department of Primary Industries and
211 Regional Development, and logistics service providers (Grain Industry Asso-
212 ciation of Western Australia, 2022). They therefore serve as useful data for
213 ground-truthing inferences drawn from the proposed model. The proposed
214 model was then applied to the island nation of Madagascar. In both cases,
215 NDVI was used to represent crop performance.

216 A geometry was set around both study regions and NDVI data derived
217 from the Moderate Resolution Imaging Spectrometer (MODIS) was down-
218 loaded in 16-day, 1 kilometre resolution (MOD13A2) via Google Earth En-
219 gine with the R interface package `rgee` (Gorelick et al., 2017; Didan, 2015;
220 Aybar et al., 2020). NDVI in this product ranges from -2,000 to 10,000. The
221 NDVI data was then masked to cropland using the Global Food-Support
222 Analysis Data Cropland Extent product (Teluguntla et al., 2014). The pe-
223 riod from January 2013 to December 2021 was chosen for Western Australia
224 to align with the comparison data available, while a longer analysis from Jan-
225 uary 2001 to December 2021 was undertaken for Madagascar. This gave a
226 final dataset of 2,968,999 and 152,459,796 observations of NDVI for Western
227 Australia and Madagascar, respectively.

228 **3. Results**

229 *3.1. Multi-field sites: Ord Irrigation Area*

230 The partial main effects of year, month, and space, and the interaction
231 effect of year and month align with patterns expected for each multi-field
232 scenario (Figures 1 and 2). Figure 2 shows an increase in the partial effect

of year on NDVI for the perennial growth scenario, an initial decrease then inter-annual variation for the perennial to annual scenario, and a relatively constant increasing trend with inter-annual variation for the spatial dynamics scenario. The partial main effect of month was similar for both the perennial to annual and spatial dynamics sites which showed a bimodal ontogeny, while the perennial growth site showed a less variable ontogeny. The partial main effect of space was more variable in the perennial growth site, while the partial effect of the interaction between month and year was more variable for the perennial to annual site, especially in later years after the change from perennial to annual crop production.

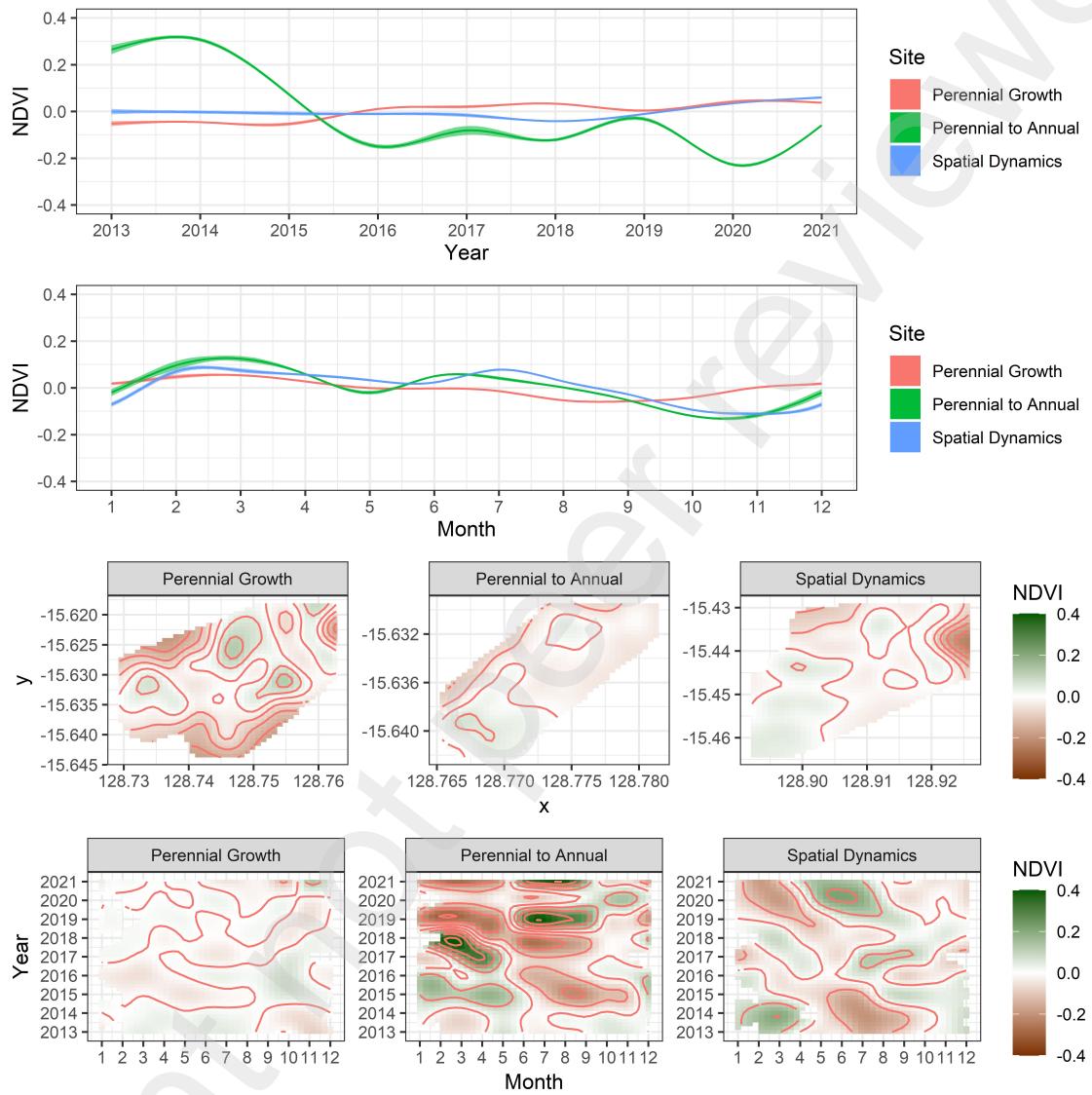


Figure 2: Partial effects for the main effect of year +/- standard error (top row), the main effect of month +/- standard error (second row), space (third row), and interaction between year and month (bottom row) for each of the perennial growth, perennial to annual, and spatial dynamics sites.

243 Figure 3 illustrates the partial effect of the interaction between space and
 244 year for the spatial dynamics as this effect was of most interest for that site.

245 It shows lesser values of NDVI in the eastern portion of the area when it
246 was being prepared for irrigated cropping in 2015, and greater values in 2018
247 once crops were established.

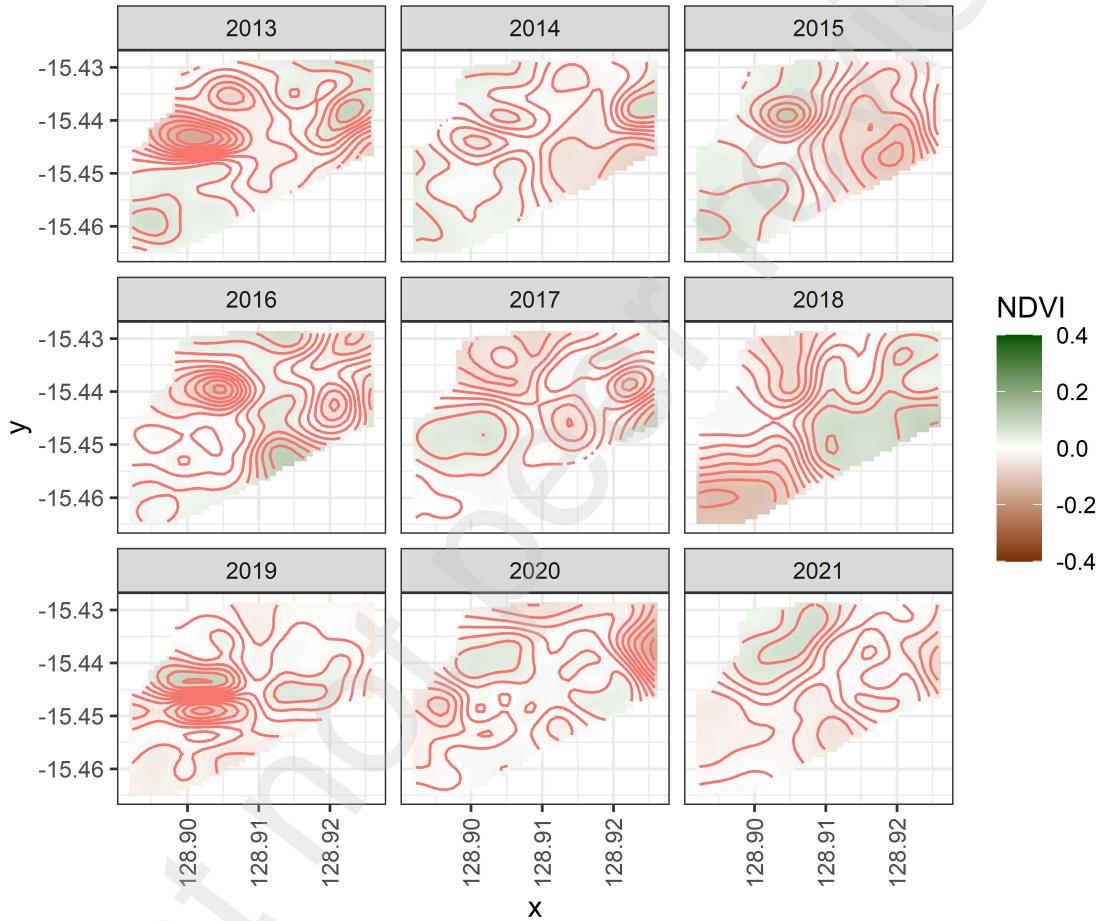


Figure 3: Partial effect of the interaction between space and year for the spatial dynamics scenario at the Ord River Irrigation Area.

248 The relative contribution of space and time effects to variation in NDVI
249 can be inferred by the loss in deviance explained when selected smooths
250 are dropped from the model. Figure 4 shows the loss in deviance explained
251 attributable to inter-annual terms (those containing 'year'). The perennial

252 to annual scenario lost the most deviance explained when these terms were
253 removed relative to other scenarios, indicating that there was greater inter-
254 annual variation in NDVI at this site, which is reflected in the partial effects
255 plots (Figure 2). Removing the interaction between year and month led to a
256 similar loss in deviance explained to removing the overall effect of year for the
257 spatial dynamics scenario, indicating that much of the inter-annual variation
258 in NDVI was caused by changes in seasonality (Table 1). The interaction
259 between space and year did not lead to much loss in deviance explained
260 across all scenarios, and was negative for the spatial dynamics scenario. This
261 suggests that spatial variation, captured in the main effect of space (Table 1),
262 was relatively constant between years. It also shows that changes in NDVI
263 in the spatial dynamics site were captured by the main effect of year and the
264 interaction between year and month, rather than the interaction between
265 year and space.

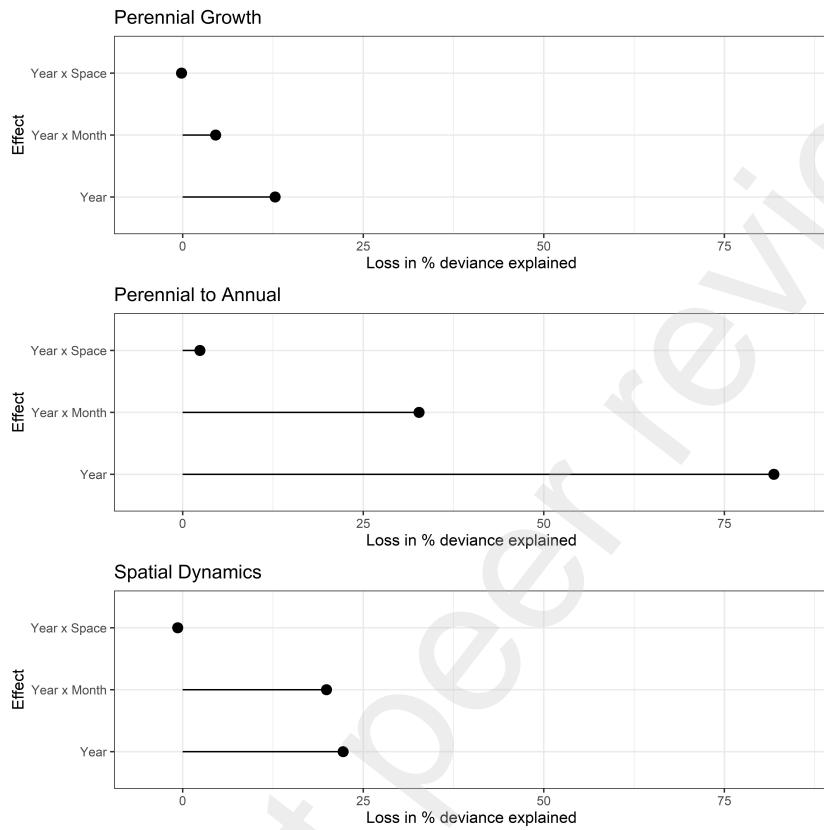


Figure 4: Loss in deviance explained attributable to the interaction between year and space, the interaction between year and month, and the overall effect of year (including interactions) for each of the three scenarios modelled at the Ord River Irrigation Area.

266 3.2. Western Australian Wheatbelt

267 The main effect of year for the Western Australian Wheatbelt shows gen-
 268 eral agreement with estimated production figures (Figure 5). A Spearman
 269 correlation test between model terms for the main effect of year and total
 270 grain production (Figure 5) gave a coefficient (ρ) of 0.73 and a p-value of
 271 0.03, indicating a significant positive correlation between the model predic-
 272 tions of NDVI and estimated grain production. Similarly, comparing the con-
 273 tribution of each port zone (Figure 5) to the partial effect of the interaction
 274 between space and year shows alignment between the on-ground estimates,

275 observations, and the results of the GAM. For example, Figure 5 shows that
276 2016 produced high NDVI levels and grain production estimates, partly at-
277 tributable to a rise in the Kwinana port zone estimate. This corresponded
278 with higher levels of NDVI in parts of the central and western Wheatbelt,
279 serving the Kwinana port zone (Figure 6). Conversely, the high production
280 year of 2021 was characterised by a greater contribution from Albany relative
281 to Kwinana port zone (Figure 5), reflected in the partial effect of space and
282 year on NDVI for 2021 (Figure 6). These inferences are further supported
283 by the observations that in 2021, the Kwinana zone had financially “the best
284 year on record, but 2016 beat it for yields” (Grain Industry Association of
285 Western Australia, 2022).

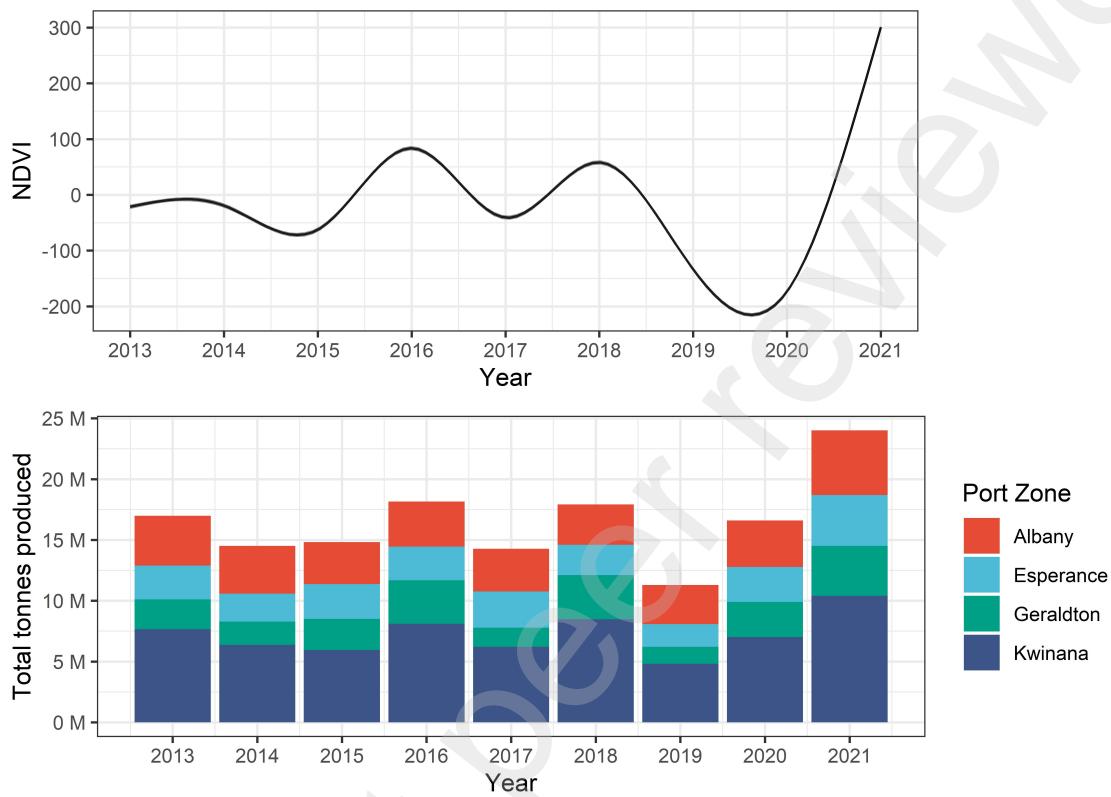


Figure 5: The main effect of year +/- standard error from a GAM trained on NDVI data for the Western Australian Wheatbelt (top) compared with estimated production volume of all grain types from the Wheatbelt by port zone and year (bottom).

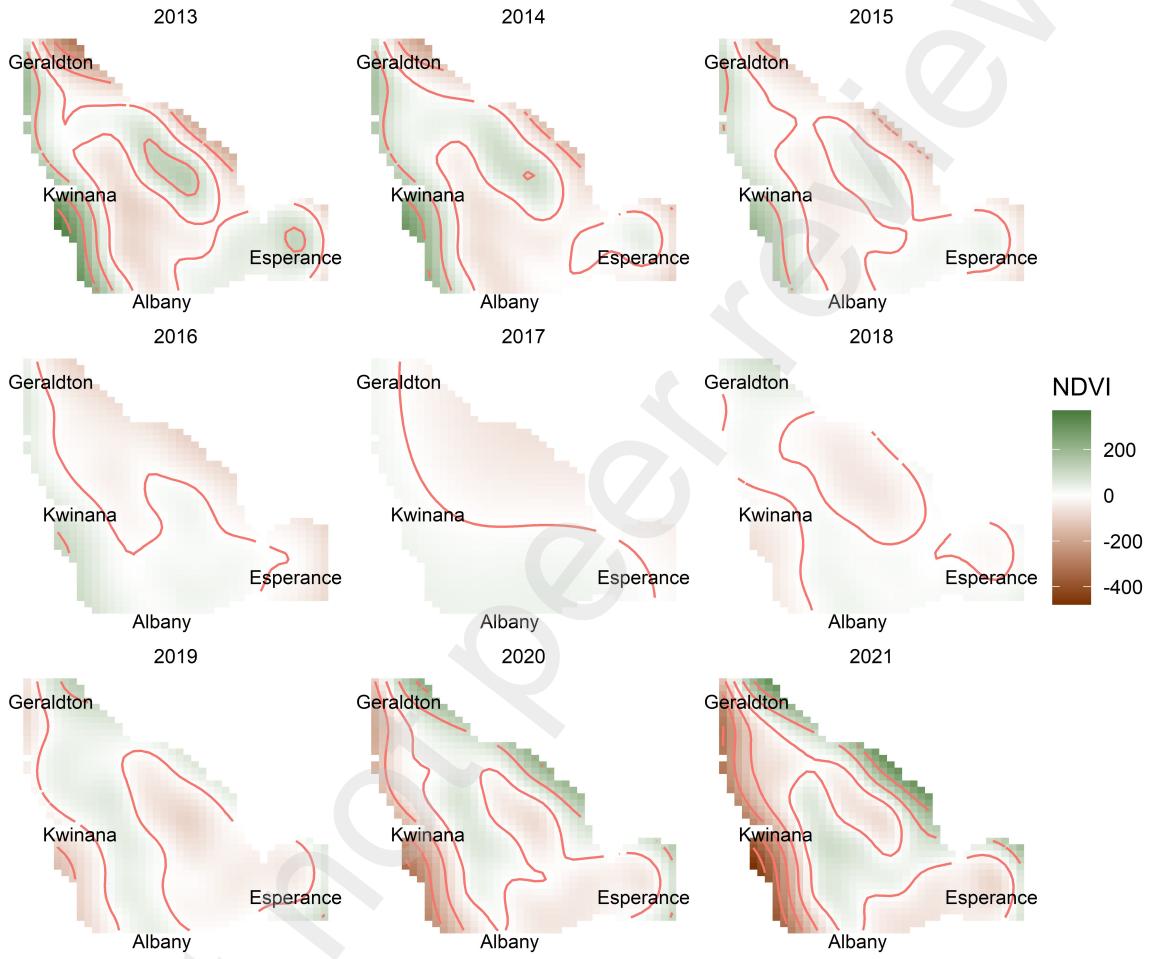


Figure 6: The partial effect of the interaction between space and year from a GAM trained on NDVI data for the Western Australian Wheatbelt, with labels denoting the approximate locations of major ports.

In addition to patterns between years and across space, extracting the partial effect of the interaction between month and year enables inference of seasonal differences. Figure 7 compares predictions of cropland NDVI

289 between the high production year of 2016 and the lower year of 2017. The
290 higher NDVI values in early 2017 before a decline matches the observation
291 made in April 2017 that “following good rains in most districts at the end
292 of January and early February, confidence was high and there was hope of a
293 repeat of the early sowing opportunities of 2016” (Grain Industry Association
294 of Western Australia, 2017a). This was followed by the May crop report
295 which noted that “the 2017 season has been a frustrating start for growers”
296 and “the trend of promising rain fronts fizzing out has continued for the last
297 month and looks to continue in the short term” (Grain Industry Association
298 of Western Australia, 2017b). Inferences drawn from the GAM therefore
299 align with on-ground grain production estimates and observations.

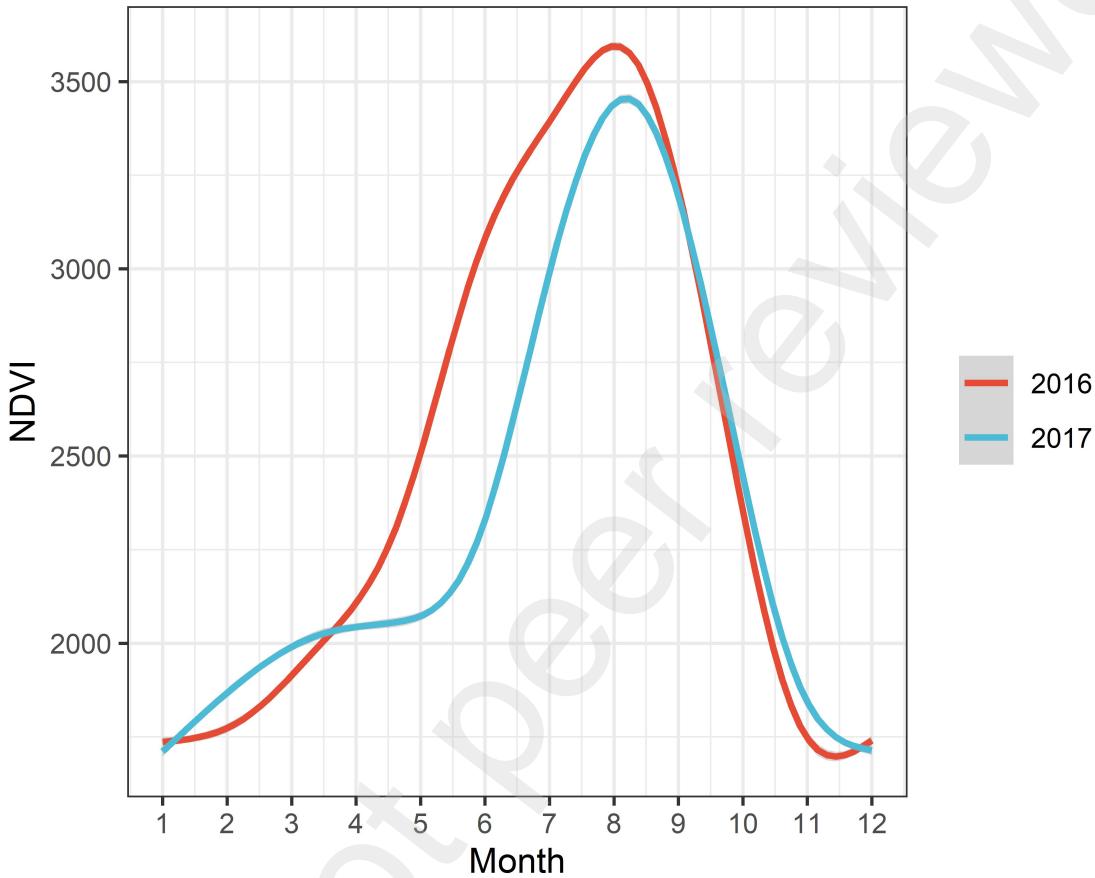


Figure 7: Predicted mean NDVI values +/- standard error aggregated across space for the Western Australian Wheatbelt in 2016 and 2017.

300 *3.3. Madagascar*

301 The partial main effect of year shows an increasing trend in cropland
 302 NDVI for Madagascar up to approximately 2014, followed by a decline (Fig-
 303 ure 8), with 2020 being the lowest year for the period analysed. Fitting the
 304 same model from 2014 to 2021, except with year as a linear term and NDVI
 305 as a log-transformed response, revealed NDVI to be declining at 2.5% per
 306 year. The main effect of month showed a crop ontogeny with a maximum
 307 around March-April, and the main effect of space showed greater NDVI val-

308 uses in the east and north of the country (Figure 8). The interaction between
 309 month and year shows some seasonal variation, notably between January and
 310 July across years.

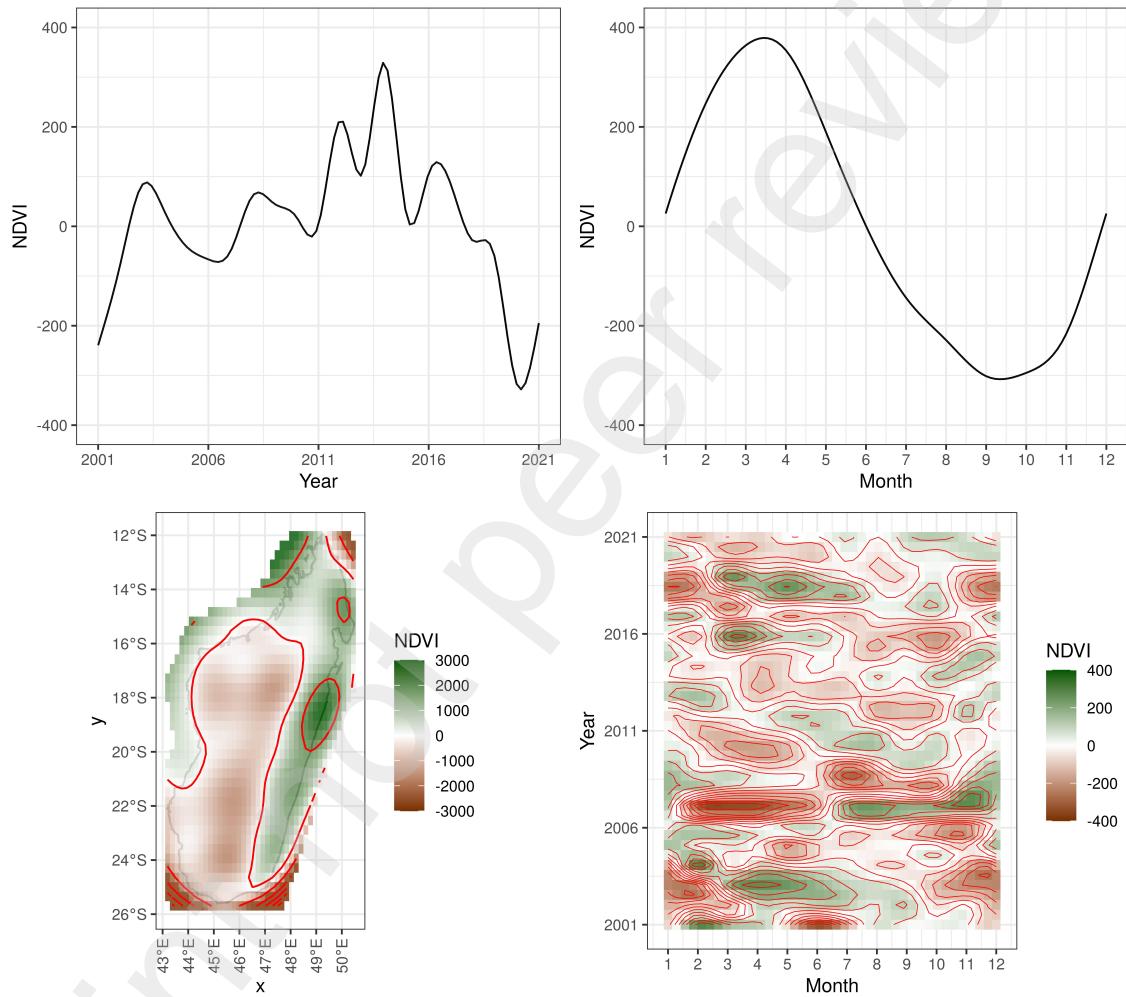


Figure 8: Partial effects of year +/- standard error (top-left), month +/- standard error (top-right), space (bottom-left), and the interaction between month and year (bottom-right).

311 The interaction between space and year for the selected years 2001, 2011,
 312 and 2021 show that 2021 was characterised by much lower values in southern

313 Madagascar relative to 2001 and 2011, which both showed relatively higher
314 values for southern Madagascar (Figure 9).

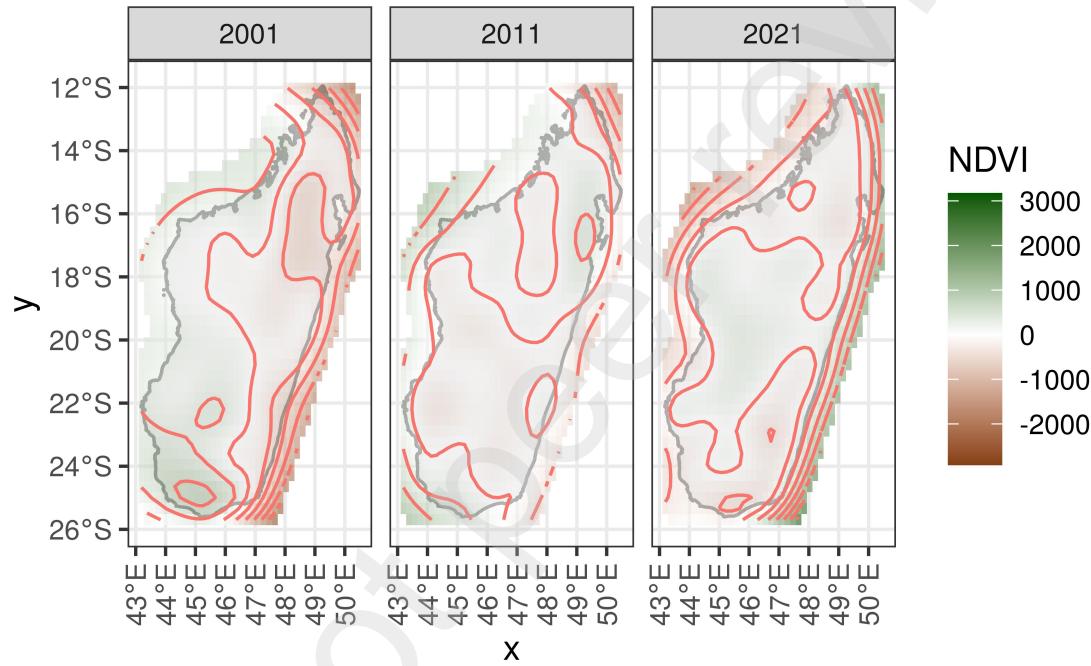


Figure 9: The partial effect of the interaction between space and year for a GAM trained on cropland NDVI data for 2001-2021 over Madagascar. The years 2001, 2011, and 2021 are plotted.

315 *3.4. Model fit and computation*

316 Table 4 shows acceptable model fit, with adjusted R^2 greater than 35%
317 for the Ord sites and the Western Australian Wheatbelt. Adjusted R^2 was
318 lower and computation time longer for the larger Madagascar dataset. Fitting

319 year as a linear rather than a smooth term for Madagascar from 2014 to 2021
320 resulted in a decrease in adjusted R².

Model	Adjusted R ² (%)	Computation time
Ord: Perennial Growth	37.7	9.5s
Ord: Perennial to Annual	91.4	3.7s
Ord: Spatial Dynamics	35.1	13.2s
Western Australian Wheat-belt	46.8	29.2s
Madagascar	23.9	35min
Madagascar- linear 2014-2021	11.8	11min

Table 4: Performance statistics for generalised additive models run on NDVI data with discretized covariates and 8 threads for three sites at the Ord River Irrigation Area and the Western Australian Wheatbelt, and on 16 threads for Madagascar.

321 There were acceptable levels of spatial autocorrelation present in all mod-
322 els fitted. Plotting a semivariogram for the perennial growth scenario (Figure
323 10) represents this. Similarly, temporal autocorrelation was corrected by fit-
324 ting an AR1 model across all sites, as demonstrated for the perennial growth
325 scenario (Figure 11).

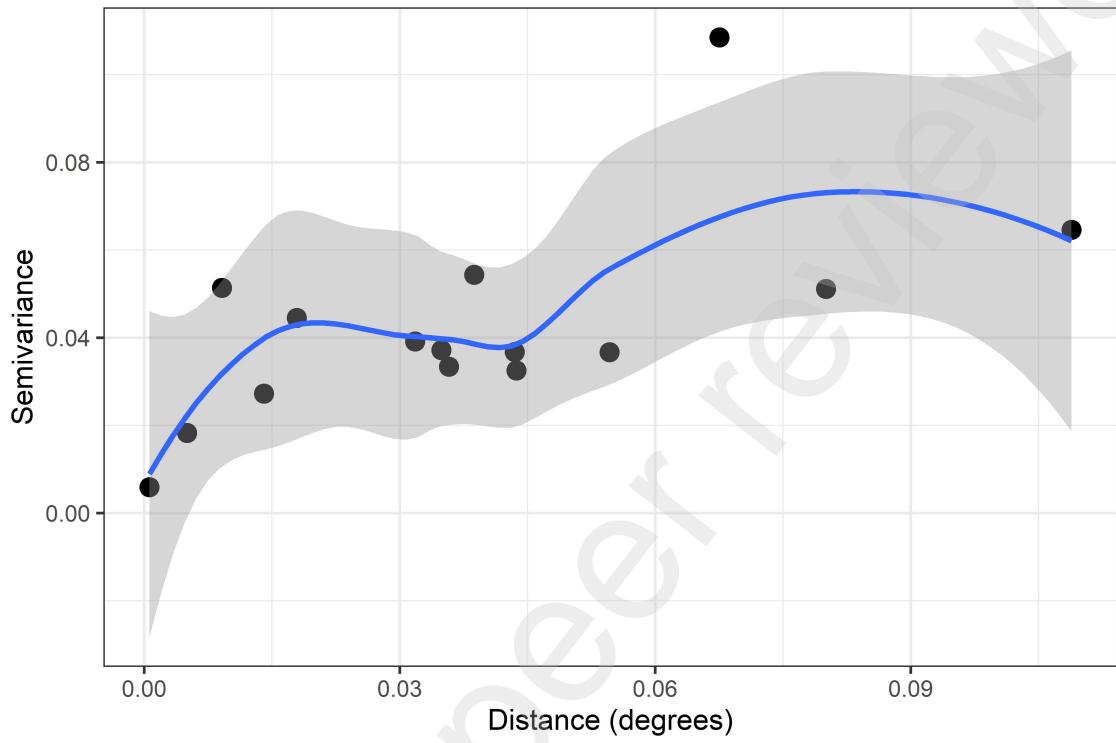


Figure 10: Variogram plot for distance for a generalised additive model fitted to NDVI from a field displaying perennial growth in the Ord River Irrigation Area. The smooth curve displayed was fitted with local polynomial regression and the ribbon shows the 95% confidence interval.

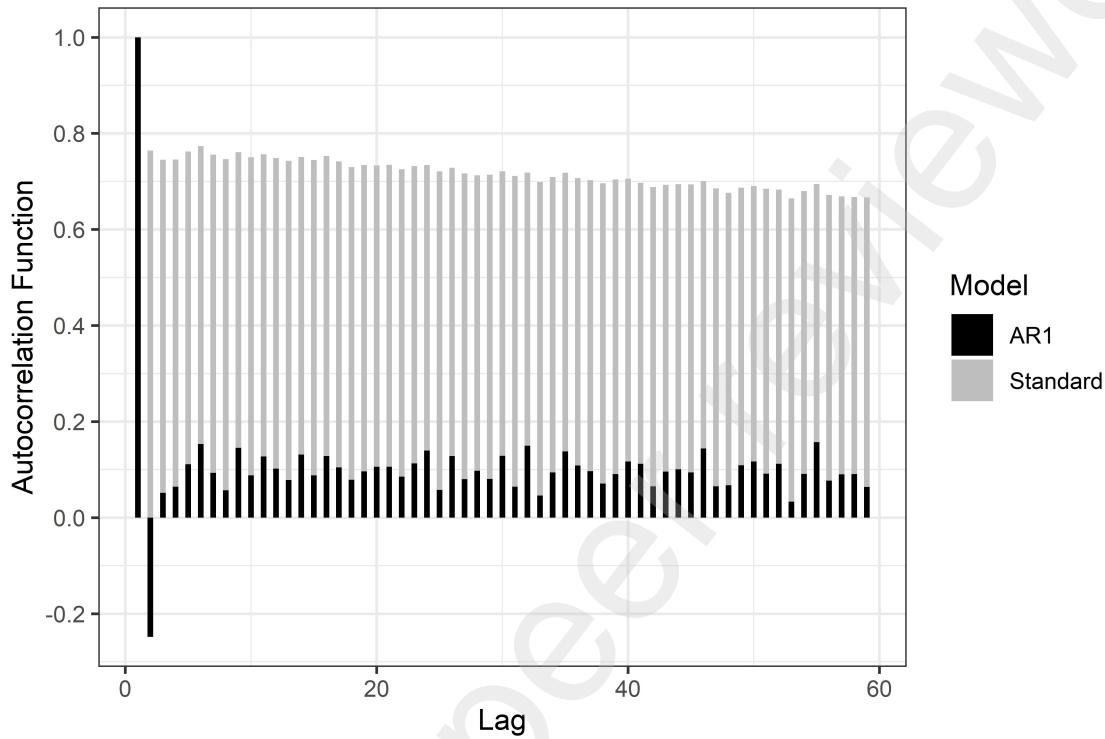


Figure 11: Aggregate autocorrelation for generalised additive model residuals from a standard initial model assuming independent residuals, overlaid with a model corrected with autoregression (AR1). Both models were fitted to NDVI from a field displaying perennial growth in the Ord River Irrigation Area.

326 4. Discussion

327 The results of our analyses demonstrate that the proposed GAM frame-
328 work can be used to infer changes in crop production across space and time
329 at a range of spatial scales. Application to several multi-field scenarios at
330 the Ord River Irrigation Area showed that the model is able to capture inter-
331 annual, seasonal, and spatial changes at farm and local scales, enabling trends
332 to be inferred (Figure 2). Inferences drawn from the same model applied over
333 the Western Australian Wheatbelt aligned with on-ground observations and
334 total production estimates (Figure 5). Patterns inferred from model out-

335 puts, with NDVI as a proxy for crop health, can therefore be interpreted as
336 associated or correlated with patterns in crop production.

337 Remote sensing assessments of well-studied regions with established mar-
338 kets, such as the Western Australian Wheatbelt, supplement other in-situ
339 estimates of crop production. However, remote sensing can play a more
340 important role where production data is scarce or sporadic, such as in Madag-
341ascar (Rigden et al., 2022). The declining trend in cropland NDVI from
342 2014 (Figure 8) raises concerns about the ability of Madagascar's agriculture
343 sector to support the population's nutritional needs, especially given the food
344 crisis since 2021 (World Food Programme, 2022). Inspection of the interac-
345 tion between year and space confirms lower NDVI values in southern Madag-
346ascar, which has also been identified by the Famine Early Warning Systems
347 (FEWS NET, 2022). However, our analysis shows that southern Madagascar
348 had relatively higher NDVI values in 2001, which also produced values below
349 0 for the partial main effect of year, indicating it was a below average pro-
350 duction year (Figures 8 & 9). Additionally, the main effect of space (Figure
351 8) does not show that southern Madagascar produces inherently lower NDVI
352 values than central Madagascar. This suggests that southern Madagascar
353 has recently become more vulnerable to drought than other regions of the
354 country, which Rakotoarison et al. (2021) found is due to a combination of
355 biophysical and socioeconomic factors. Overall, our analysis confirms the
356 greater vulnerability of southern Madagascar to crop production shocks, in
357 addition to a declining trend in cropland NDVI since 2014. This finding gives
358 evidence for much lower agricultural production in southern Madagascar in
359 recent years, and agricultural institutions should prioritise research activities
360 that stabilise the food supply in this region.

361 The proposed GAM framework enables inferences to be drawn across
362 space and time from a single model. The flexibility of the smooth terms
363 enables patterns to be observed which may not be identified in more para-
364 metric approaches. For example, the identification of bimodal ontogeny in

365 the partial main effect of month for the Ord Irrigation Area scenarios (Figure 2) is unlikely to be identified by ontogeny or trend detection approaches
366 which necessitate a single maxima (Younes et al., 2020). This feature of the
367 framework enables detection of more than one type of pasture or crop in a
368 farming system, and also contributes to detecting changes in seasonality.

370 Advances in computation of GAMs and access to improved computational
371 infrastructure has enabled this statistical approach to be applied to
372 very large remote sensing datasets. Without the discretised covariate ap-
373 proach, the computational requirements of the proposed model would be
374 prohibitively costly (Li and Wood, 2020; Wood et al., 2017). Similarly, the
375 ability to fit an autoregressive error overcomes the limitation of temporal
376 autocorrelation (Wood et al., 2017). However, the abundance of data as-
377 sociated with remote sensing introduces some limitations and challenges for
378 model fit and interpretation. For example, the very large dataset size ef-
379 fectively voids the hypothesis testing procedure adapted to GAMs (Wood,
380 2013). Table 4 also shows that the larger datasets for Ord: Spatial Dynamics
381 and Perennial Growth, and Madagascar, produced lower model fit statistics.
382 This could be addressed through a more flexible spatial term by raising the
383 maximum effective degrees of freedom, k , especially if a model is intended
384 for prediction, although this increases computational costs. Further develop-
385 ment of approaches that account for spatio-temporal variation in very large
386 data with computational efficiency will enhance the application of statistical
387 modelling approaches to remote sensing.

388 The proposed framework is not limited to NDVI as a response, or time
389 and space as exclusive predictors. Users may consider a response that more
390 directly represents crop production, such as predicted yield if it is available.
391 Alternatively, users may be interested in modelling another variable asso-
392 ciated with crop growth, such as transpiration or evapotranspiration which
393 could allow inference of trends in water use efficiency (Hsiao et al., 2007).
394 In this instance, other predictors could be fitted, such as soil moisture or

395 rainfall. A rainfall lag, drought, or climate index could also be fitted to
396 the model we have proposed for Madagascar which may assist in differen-
397 tiating trends attributable to climatic factors from other factors, such as
398 socioeconomic effects (Rakotoarison et al., 2021). The proposed framework
399 is therefore adaptable, though should be built around the space and time
400 components.

401 5. Conclusion

402 A framework using GAMs, as a semi-parametric statistical approach, en-
403 abled the effects of space and time on crop production to be decomposed and
404 inferences to be drawn. This can inform important food security analyses,
405 such as the detection of a decreasing cropland NDVI trend in Madagascar.
406 The proposed framework is adaptable and can be applied to numerous agri-
407 cultural research questions. However, ongoing improvements in managing
408 and modelling large spatio-temporal datasets are likely to improve its appli-
409 cation.

410 6. Acknowledgements

411 This research was undertaken while supported by the Australian National
412 University (ANU) University Research Scholarship and a Commonwealth Sci-
413 entific and Industrial Research Organisation (CSIRO) and ANU Digital Agri-
414 culture Supplementary Scholarship through the Centre for Entrepreneurial
415 Agri-Technology.

416 References

- 417 Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M.,
418 Ghemawat, S., Irving, G., Isard, M., 2016. TensorFlow: A System for
419 Large-Scale Machine Learning, in: 12th USENIX symposium on operating
420 systems design and implementation (OSDI 16), pp. 265–283.

- 421 Aybar, C., Wu, Q., Bautista, L., Yali, R., Barja, A., 2020. rgee: An R
422 package for interacting with Google Earth Engine. Journal of Open Source
423 Software 5, 2272.
- 424 Cassman, K.G., Grassini, P., 2013. Can there be a green revolution in
425 sub-saharan africa without large expansion of irrigated crop production?
426 Global Food Security 2, 203–209.
- 427 CGIAR, 2019. CGIAR Platform for Big Data in Agriculture Annual Report
428 2018. Report.
- 429 Chen, K., O’Leary, R.A., Evans, F.H., 2019. A simple and parsimonious
430 generalised additive model for predicting wheat yield in a decision support
431 tool. Agricultural Systems 173, 140–150. URL: [https://www.scienc](https://www.sciencedirect.com/science/article/pii/S0308521X18308667)
432 [edirect.com/science/article/pii/S0308521X18308667](https://www.sciencedirect.com/science/article/pii/S0308521X18308667), doi:<https://doi.org/10.1016/j.agrsy.2019.02.009>.
- 434 De Rosa, D., Basso, B., Fasiolo, M., Friedl, J., Fulkerson, B., Grace, P.R.,
435 Rowlings, D.W., 2021. Predicting pasture biomass using a statistical model
436 and machine learning algorithm implemented with remotely sensed im-
437 agery. Computers and Electronics in Agriculture 180, 105880. URL:
438 <https://www.sciencedirect.com/science/article/pii/S0168169920330854>, doi:<https://doi.org/10.1016/j.compag.2020.105880>.
- 440 Dhu, T., Dunn, B., Lewis, B., Lymburner, L., Mueller, N., Telfer, E., Lewis,
441 A., McIntyre, A., Minchin, S., Phillips, C., 2017. Digital Earth Australia
442 – unlocking new value from earth observation data. Big Earth Data 1,
443 64–74. URL: <https://doi.org/10.1080/20964471.2017.1402490>,
444 doi:[10.1080/20964471.2017.1402490](https://doi.org/10.1080/20964471.2017.1402490).
- 445 Didan, K., 2015. MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3
446 Global 1km SIN Grid V006. Report. URL: <https://doi.org/10.5067/MODIS/MOD13A2.006>.

- 448 Dossou-Yovo, E.R., Vandamme, E., Dieng, I., Johnson, J.M., Saito, K., 2020.
- 449 Decomposing rice yield gaps into efficiency, resource and technology yield
- 450 gaps in sub-Saharan Africa. *Field Crops Research* 258, 107963.
- 451 Edmondson, M., 2019. *googleComputeEngineR*: R Interface with Google
- 452 Compute Engine. Report. URL: <https://CRAN.R-project.org/package=googleComputeEngineR>.
- 453
- 454 Fasiolo, M., Nedellec, R., Goude, Y., Capezza, C., Wood, S.N., Fasiolo, M.M., 2020a. Package ‘mgcviz’. Visualisations for generalised additive models. Available from URL: <https://cran.r-project.org/web/packages/mgcviz/index.html>.
- 455
- 456
- 457
- 458 Fasiolo, M., Nedellec, R., Goude, Y., Wood, S.N., 2020b. Scalable Visualization Methods for Modern Generalized Additive Models. *Journal of Computational and Graphical Statistics* 29, 78–86. URL: <https://doi.org/10.1080/10618600.2019.1629942>, doi:10.1080/10618600.2019.1629942.
- 459
- 460
- 461
- 462 FEWS NET, 2022. Madagascar. Technical Report. URL: <https://fews.net/southern-africa/madagascar>.
- 463
- 464 Fewster, R.M., Buckland, S.T., Siriwardena, G.M., Baillie, S.R., Wilson, J.D., 2000. Analysis of population trends for farmland birds using generalized additive models. *Ecology* 81, 1970–1984. URL: <https://doi.org/10.2307/177286>, doi:[https://doi.org/10.1890/0012-9658\(2000\)081\[1970:AOPTF\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[1970:AOPTF]2.0.CO;2).
- 465
- 466
- 467
- 468
- 469 Fischer, R., Byerlee, D., Edmeades, G., 2014. Crop yields and global food security: will yield increase continue to feed the world? ACIAR Monograph Series, Australian Centre for International Agricultural Research, Canberra, ACT.
- 470
- 471
- 472
- 473 Fletcher, A., Weeks, C., Lawes, R., 2016. Why are WA farmers sowing earlier or dry sowing. Report. URL: <https://grdc.com.au/resources-and-pu>
- 474

- 475 [blications/grdc-update-papers/tabccontent/grdc-update-papers
476 /2016/03/why-are-wa-farmers-sowing-earlier-or-dry-sowing](https://www.grdc.update-papers.com.au/2016/03/why-are-wa-farmers-sowing-earlier-or-dry-sowing).
- 477 Forkel, M., Carvalhais, N., Verbesselt, J., Mahecha, M.D., Neigh, C.S., Re-
478 ichstein, M., 2013. Trend Change Detection in NDVI Time Series: Effects
479 of Inter-Annual Variability and Methodology. *Remote Sensing* 5, 2113–
480 2144. URL: <https://www.mdpi.com/2072-4292/5/5/2113>.
- 481 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore,
482 R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for
483 everyone. *Remote Sensing of Environment* 202, 18–27. URL: <https://www.sciencedirect.com/science/article/pii/S0034425717302900>,
484 doi:<https://doi.org/10.1016/j.rse.2017.06.031>.
- 485 Grain Industry Association of Western Australia, 2017a. GIWA Crop Report-
486 April 2017. Technical Report. URL: <https://1qk7mll6sku1bnrjbjn3n81e-wpengine.netdna-ssl.com/wp-content/uploads/2021/01/GIWACropReportApril2017-1.pdf>.
- 487 Grain Industry Association of Western Australia, 2017b. GIWA Crop Report-
488 May 2017. Technical Report. URL: <https://1qk7mll6sku1bnrjbjn3n81e-wpengine.netdna-ssl.com/wp-content/uploads/2021/01/GIWACropReportMay2017-1.pdf>.
- 489 Grain Industry Association of Western Australia, 2022. GIWA Crop Report-
490 February 2022. Technical Report. URL: <https://www.giwa.org.au/wa-crop-reports/2021-season/giwa-crop-report-february-2022/>.
- 491 Hastie, T.J., Tibshirani, R.J., 2017. Generalized additive models. Routledge.
- 492 Holloway, J., Mengersen, K., 2018. Statistical Machine Learning Methods
493 and Remote Sensing for Sustainable Development Goals: A Review. *Re-
494 mote Sensing* 10, 1365. URL: <https://www.mdpi.com/2072-4292/10/9/1365>.

- 502 Hsiao, T.C., Steduto, P., Fereres, E., 2007. A systematic and quantitative
503 approach to improve water use efficiency in agriculture. Irrigation science
504 25, 209–231.
- 505 Krause, C., Dunn, B., Bishop-Taylor, R., Adams, C., Burton, C., Alger, M.,
506 Chua, S., Phillips, C., Newey, V., Kouzoubov, K., 2021. Digital Earth
507 Australia notebooks and tools repository. Geoscience Australia .
- 508 Li, Z., Wood, S.N., 2020. Faster model matrix crossproducts for large gener-
509 alized linear models with discretized covariates. Statistics and Computing
510 30, 19–25. URL: <https://doi.org/10.1007/s11222-019-09864-2>,
511 doi:10.1007/s11222-019-09864-2.
- 512 Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable
513 satellite-based crop yield mapper. Remote Sensing of Environment 164,
514 324–333. URL: <https://www.sciencedirect.com/science/article/pii/S0034425715001637>, doi:<https://doi.org/10.1016/j.rse.2015.04.021>.
- 517 Ouattara, B., Forkuor, G., Zoungrana, B.J.B., Dimobe, K., Danumah, J.,
518 Saley, B., Tondoh, J.E., 2020. Crops monitoring and yield estimation using
519 sentinel products in semi-arid smallholder irrigation schemes. International
520 Journal of Remote Sensing 41, 6527–6549. URL: <https://doi.org/10.1080/01431161.2020.1739355>, doi:<https://doi.org/10.1080/01431161.2020.1739355>.
- 523 Paciorek, C.J., 2003. Nonstationary Gaussian processes for regression and
524 spatial modelling. Thesis.
- 525 Pretty, J.N., 1997. The sustainable intensification of agriculture. Natural
526 Resources Forum 21, 247–256. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1477-8947.1997.tb00699.x>, doi:10.1111/j.1477-8947.1997.tb00699.x.

- 529 R Core Team, 2022. R: A Language and Environment for Statistical
530 Computing. Report. R Foundation for Statistical Computing. URL:
531 <https://www.R-project.org/>.
- 532 Rakotoarison, T.R., Hajalalaina, A.R., Raonivelox, A., Raherininirina, A., Zo-
533 jaona, R.T., 2021. Spatial Analysis of Risks and Vulnerabilities to Major
534 Hazards in Madagascar Using the Multi-Criteria Method Based on the An-
535 alytical Hierarchy Process (AHP). Journal of Geoscience and Environment
536 Protection 9, 15–24.
- 537 Reardon, T., Crawford, E.W., Kelly, V.A., Diagana, B.N., 1995. Promot-
538 ing farm investment for sustainable intensification of African agriculture.
539 Report.
- 540 Rigden, A.J., Golden, C., Huybers, P., 2022. Retrospective Predictions of
541 Rice and Other Crop Production in Madagascar Using Soil Moisture and
542 an NDVI-Based Calendar from 2010–2017. Remote Sensing 14, 1223.
- 543 Rockström, J., Williams, J., Daily, G., Noble, A., Matthews, N., Gordon,
544 L., Wetterstrand, H., DeClerck, F., Shah, M., Steduto, P., de Fraiture, C.,
545 Hatibu, N., Unver, O., Bird, J., Sibanda, L., Smith, J., 2017. Sustainable
546 intensification of agriculture for human prosperity and global sustainabil-
547 ity. Ambio 46, 4–17. URL: <https://doi.org/10.1007/s13280-016-0793-6>.
548
- 549 Rouse, J., Haas, R., Schell, J., 1973. Monitoring vegetation systems in the
550 great plains with ERTS, in: Third ERTS Symposium.
- 551 Royston, P., Sauerbrei, W., 2007. Multivariable modeling with cubic regres-
552 sion splines: a principled approach. The Stata Journal 7, 45–70.
- 553 Simpson, G.L., 2018. Modelling Palaeoecological Time Series Using Gen-
554 eralised Additive Models. Frontiers in Ecology and Evolution 6. URL:

- 555 <https://www.frontiersin.org/article/10.3389/fenv.2018.00149>,
556 doi:10.3389/fenv.2018.00149.
- 557 Stephens, D., Lyons, T., 1998. Variability and trends in sowing dates across
558 the Australian wheatbelt. Australian Journal of Agricultural Research 49,
559 1111–1118.
- 560 Teluguntla, P., Thenkabail, P., Xiong, J. and Gumma, M., Giri, C., Milesi, C.,
561 Ozdogan, M., Congalton, R., Tilton, J., Sankey, T., Massey, R., Phalke,
562 A., Yadav, K., 2014. Global Food Security Support Analysis Data at
563 Nominal 1 km (GFSAD1km) Derived from Remote Sensing in Support of
564 Food Security in the Twenty-First Century: Current Achievements and
565 Future Possibilities. Report.
- 566 Van Rij, J., Wieling, M., Baayen, R.H., van Rijn, D., 2015. itsadug: Inter-
567 preting time series and autocorrelated data using GAMMs .
- 568 Venables, W.N., Dichmont, C.M., 2004. GLMs, GAMs and GLMMs: an
569 overview of theory for applications in fisheries research. Fisheries Research
570 70, 319–337. URL: <https://www.sciencedirect.com/science/article/pii/S0165783604001754>, doi:<https://doi.org/10.1016/j.fishres.2004.08.011>.
- 573 Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting
574 trend and seasonal changes in satellite image time series. Remote sensing
575 of Environment 114, 106–115.
- 576 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François,
577 R., Grolemund, G., Hayes, A., Henry, L., Hester, J., 2019. Welcome to
578 the Tidyverse. Journal of Open Source Software 4, 1686. doi:<https://doi.org/10.21105/joss.01686>.
- 580 Wikle, C.K., Zammit-Mangion, A., Cressie, N., 2019. Spatio-temporal Statistics
581 with R. Chapman and Hall/CRC.

- 582 Wood, S., Wood, M.S., 2015. Package ‘mgcv’. R package version 1, 729.
- 583 Wood, S.N., 2006. Generalized additive models: an introduction with R.
584 chapman and hall/CRC.
- 585 Wood, S.N., 2013. On p-values for smooth components of an extended gen-
586 eralized additive model. *Biometrika* 100, 221–228.
- 587 Wood, S.N., Goude, Y., Shaw, S., 2015. Generalized additive models for
588 large data sets. *Journal of the Royal Statistical Society: Series C (Applied
589 Statistics)* 64, 139–155. URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12068>, doi:<https://doi.org/10.1111/rssc.12068>.
- 591
- 592 Wood, S.N., Li, Z., Shaddick, G., Augustin, N.H., 2017. Generalized Ad-
593 ditive Models for Gigadata: Modeling the U.K. Black Smoke Network
594 Daily Data. *Journal of the American Statistical Association* 112, 1199–
595 1210. URL: <https://doi.org/10.1080/01621459.2016.1195744>,
596 doi:[10.1080/01621459.2016.1195744](https://doi.org/10.1080/01621459.2016.1195744).
- 597 World Food Programme, 2022. Southern Madagascar emergency. Report.
598 URL: <https://www.wfp.org/emergencies/southern-madagascar-em->
599 [ergency](#).
- 600 Yee, T.W., Mitchell, N.D., 1991. Generalized additive models in plant
601 ecology. *Journal of Vegetation Science* 2, 587–602. URL: <https://onlinelibrary.wiley.com/doi/abs/10.2307/3236170>, doi:<https://doi.org/10.2307/3236170>.
- 603
- 604 Younes, N., Northfield, T.D., Joyce, K.E., Maier, S.W., Duke, N.C., Lym-
605 burner, L., 2020. A novel approach to modelling mangrove phenology from
606 satellite images: A case study from northern Australia. *Remote Sensing*
607 12, 4008.