The Earth Institute
Columbia University

The impacts of climate change on coffee: trouble brewing

James Rising, Jeffrey Sachs,
Tim Foreman, John Simmons, Manuel Brahm
The Earth Institute
Columbia University

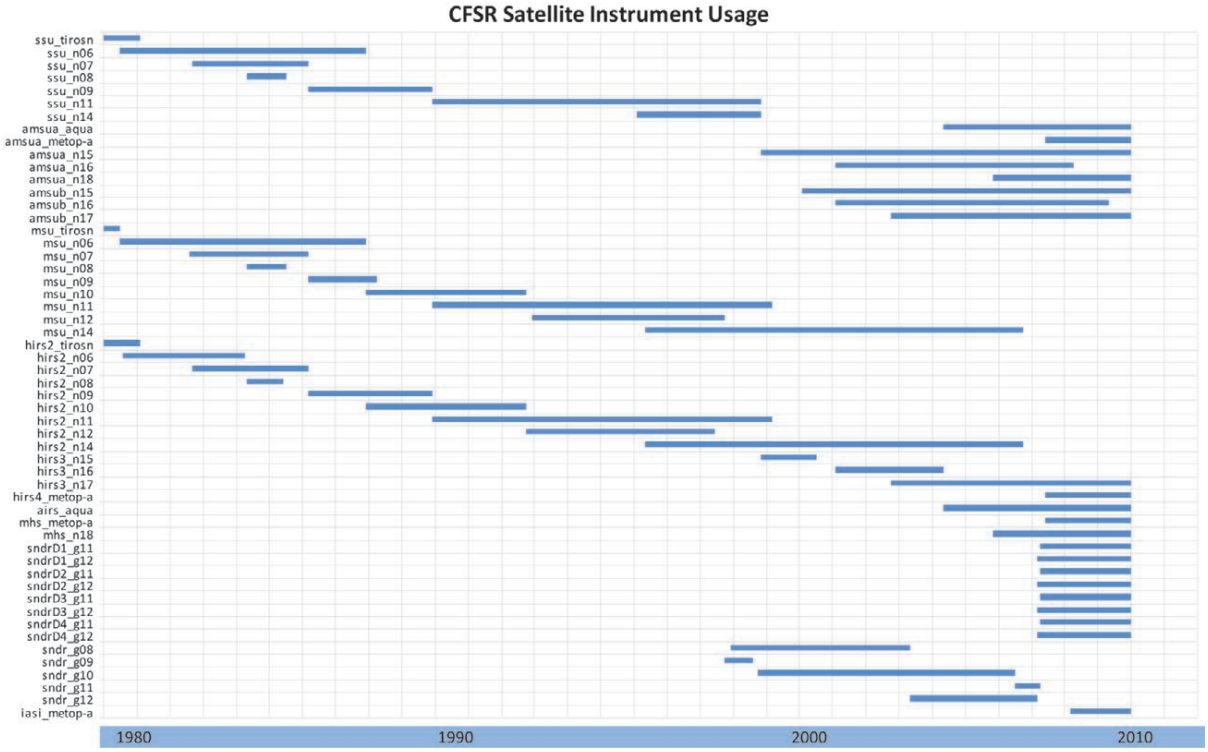
November 7, 2016

Contents

Supplemental material

0.1 Climate Forecast System Reanalysis

The Climate Forecast System Reanalysis is a global weather product constructed by NOAA (Saha et al., 2010). CFSR merges the overlapping ranges of satellite products, as they are available across years:



CFSR combines both conventional and satellite data from the following sources:

Conventional: Radiosondes and Pibals, AMMA special observations, Aircraft and ACARS data, Surface observations, PAOBS, SATOB observations, SSM/I ocean surface wind speed, Scatterometer winds

Satellite-radiance: TOVS radiances, Recalibrated MSU radiances, ATOVS radiances, GEOS radiances, Aqua AIRS, AMSU-A, and AMSR-E data, MetOp IASI, AMSU-A, and MHS data, CHAMP/-COSMIC GPS radio occultation data.

0.2 Brazil case study

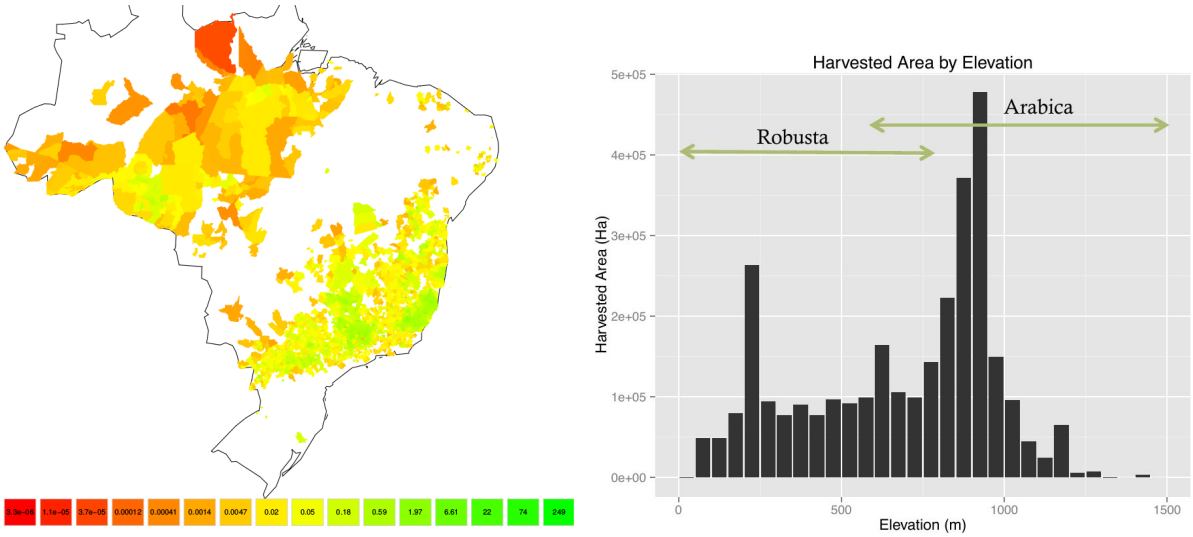


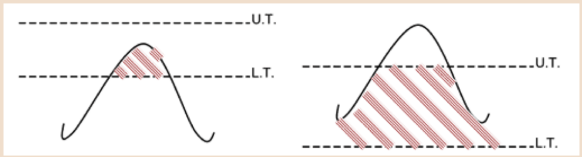
Figure 1: Brazil dataset across space and elevation. **Left:** Density of coffee production, as the average production divided by municipality area. Regions in green account for the majority of production. Most production occurs in the south, however there are coffee producing regions also in the southern Amazon. **Right:** Distribution of coffee producing area, displayed across the average elevation of each municipality. The greatest extent of coffee production occurs in municipalities with around 900 m of elevation, but coffee is also produced in municipalities with a much lower elevation, including a peak around 200 m. The range of typical elevations for growing Arabica and Robusta are shown above the histogram.

0.2.1 An empirical model of production

Using the IBGE Brazilian coffee production estimates, combined with high resolution weather from the CFSR reanalysis product, we estimate a physically-based statistical model of coffee production. The model predicts yields using a nonlinear relationship with temperature and precipitation. We base our model on Schlenker and Roberts (2009), and divide GDDs into three groups: beneficial growing degree-days between 0°C and 33°C, killing degree-days above 33°C, and frost degree days below 0°C. We also use the average minimum temperature, which appears to be more significant than frost degrees. This kind of statistical relationship is based on the biological response of coffee to temperature, but puts a "black box" around farmer responses and ecosystem and pest dynamics. If farmers are providing sufficient irrigation and shade to coffee plants, the effect of high temperatures will be mitigated beyond what biological models suggest on their own.

Calculating growing degree-days

Growing degree-days (GDDs) are calculated using a continuous sinusoidal fit to minimum and maximum daily temperatures, as shown below:



Calculations for growing degree-days and killing degree-days. Any temperatures above a given lower threshold (L.T.) are included, up to a maximum of an upper threshold (U.T.). As temperatures shift over the course of a day, fractional growing degree-days are accumulated.

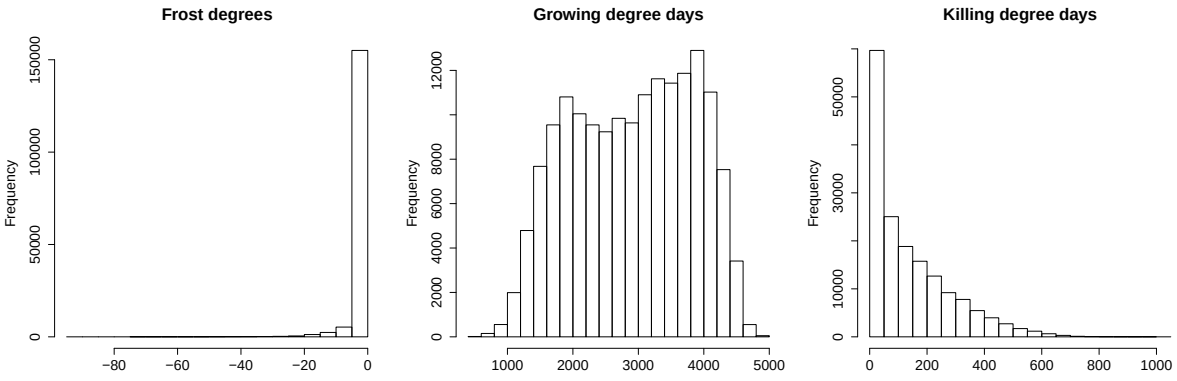


Figure 2: Histograms displaying the number of growing seasons with a given number of frost degree-days, growing degree-days, and killing degree-days. The exponential decays in frost and killing degree days are useful for capturing the impact of extreme events. The broad range of growing degree-days represented in the center histogram allows for accurate estimates of the coffee growth response.

We also include precipitation, as the total accumulated precipitation over the six months before harvest. Precipitation is included as a quadratic, to capture the expectation that both too little precipitation and too much precipitation are harmfully impact yields.

0.2.2 Optimal temperature range

Guzmán Martínez et al. (1999) suggest that 10°C is the appropriate base temperature for calculating GDDs for coffee. We explore a large range of minimum and maximum temperatures for GDDs, seeking the limits that provide the greatest predictive capacity. See Appendix 0.5.2 for predictive capacity of a range of possible limits. We find that a minimum temperature of 0°C and a maximum temperature of 33°C for beneficial GDDs is optimal. This means not only that all days over 0°C are estimated as beneficial, but that higher temperatures up to 33°C are progressively more beneficial. A day above 33°C is not immediately detrimental, but it has a progressively smaller benefit until it becomes negative, and we find that temperatures over about 35°C are detrimental in Brazil.

0.2.3 Predictive periods

Coffee production is very sensitive to weather during flowering, and the period during which we correlate weather with yields is important. To determine the optimal span of weather for predicting yields, we try out many combinations of starting and ending months. The harvesting period in Brazil ends in September, so we consider months starting with October to predict the yield in the next year. The coefficients of models for each of these periods are shown in figure 3.

A few features are important in these results. In the top graph displaying coefficient values, areas in the upper-left are gray, denoting that models that use only the months shortly preceding harvest do not produce significant results. Second, we expect the effect of GDDs to be positive, KDDs negative, the linear component of precipitation (`precip`) to be positive, and the quadratic component of it (`precip2`) to be negative. This is confirmed for most date ranges, and we want to avoid regions that misestimate these values due to noisy or minor effects. Finally, the t-values figures show the confidence in these values, and are a measure of the statistical significance of the model as a whole. These values generally decrease as the starting month becomes later.

Figure 4 shows the combined t-values for the GDD and KDD coefficients. The highest t-value is for GDD and KDD values calculated just for January and February. The probably reflects a highly sensitive period for the berry production. Nearly as high, and covering a six-month span, is December through May. We will use this as our span for calculating weather impacts.

0.2.4 Econometric model

The form of the statistical model is,

$$\log y_{it} = \alpha_i + \gamma g_{it} + \kappa k_{it} + \mu m_{it} + \pi p_{it} + \psi p_{it}^2 + P_{3,s(i)}(t) + \epsilon_{it}$$

Above and in the other models below, the observation variables and their corresponding effect estimating coefficients are:

	Var.	Coeff.
Growing degree-days	g_{it}	γ
Killing degree-days	k_{it}	κ
Average minimum temperature	m_{it}	μ
Total precipitation (linear)	p_{it}	π
Total precipitation (quadratic)	p_{it}^2	ψ

where i indexes municipalities, t the years, and $P_{3,s(i)}(t)$ is a state-specific cubic trend to capture shifting productive capacity. We aggregate weather from December to May, and use 0°C to 33°C as the limits for computing growing degree-days.

Interpreting regression tables

Many of the results in this chapter are in the form of multiple regression tables. Each regression is of the form,

$$y_i = \alpha + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + \epsilon_i$$

which describes the relationship between a dependent variable, y_i, taking different values for each ith observation, and a linear combination of independent variables, $x_{1,i}, \ldots, x_{k_i}$. The ϵ_i term represents the remaining error that cannot be explained by the model. In addition, these models use "fixed-effects", which are parameters unique to each region, so that the model is effectively estimated by

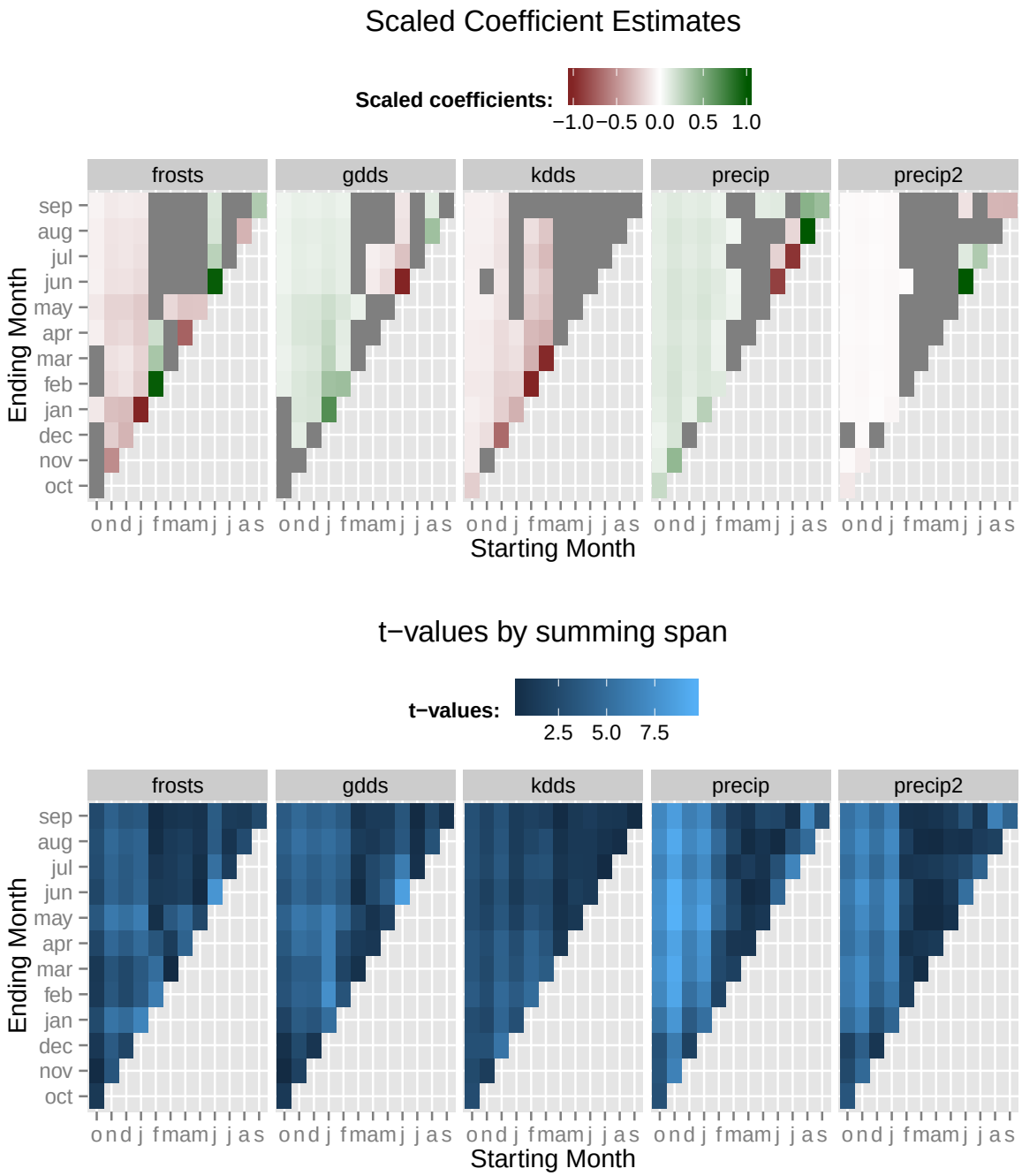


Figure 3: Coefficients from estimating models with different month spans, and the t-values intervals associated with each coefficient. The top 118 municipalities in harvest density were used.

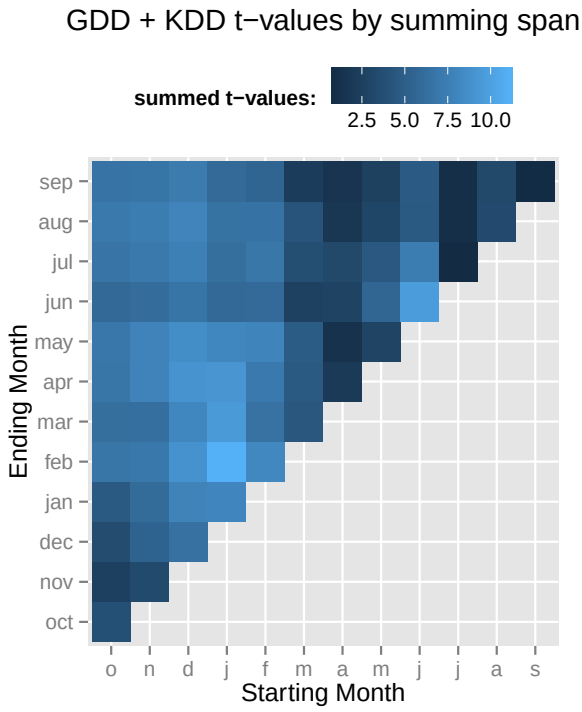


Figure 4: The sum of t-values across the GDD and KDD coefficients, for identifying the most effective range.

considering the effects of changes in the independent variables, rather underlying static differences between them.

The regression tables are mean to be read in columns. The first column specifies the variable for which an effect is reported, and the model columns specify the size of that effect. If a coefficient estimated is 10, that means that the dependent variable increases by 10 for every unit the independent variable increases.

The numbers directly below each effect and reported in parentheses are the values 'standard errors', a measure of the uncertainty of that value. If the standard error is less than half of the value, then there is 95% confidence that the sign of the coefficient in question is correct. This corresponds to the statistical significance of the estimate, and is denoted by asterisks (***).

The results are shown below as a table of statistical coefficients. Table 1 displays the results across all municipalities, and 2 is for the 118 municipalities with the greatest density of coffee harvesting.

		Dependent variable:	
	Means	Log Yields	Harvested Hectares
		(1)	(2)
GDDs / 1000	2.946	0.152***	72.869
	(0.931)	(0.050)	(124.246)
KDDs / 1000	0.149	−2.806***	−2,197.369***
	(0.146)	(0.342)	(555.055)
Avg. Min.	0.944	−0.091***	−25.0
	(3.499)	(0.018)	(34.0)
Precip. (m)	1.421	0.347***	−9.587
	(0.719)	(0.028)	(64.092)
Precip.^{2} (m)	2.538	−0.366***	−8.520
	(2.439)	(0.036)	(84.618)
State cubic trends		Yes	Yes
Observations		43,165	43,185
R^2		0.383	0.655
Adjusted R^2		0.343	0.633
Residual Std. Error		0.535 (df = 40542)	4,300.446 (df = 40561)
Note:		*p<0.1; **p<0.05; ***p<0.01	

Table 1: Estimates for statistical models relating growing degree-days, killing degree-days, average minimum temperature, and precipitation to the logarithm of yields, and to harvested area, for all municipalities. Stars (***) represent statistical significance levels, showing that most coefficients appear to have a relationship with production outputs.

0.2.5 Multilevel Brazil model

Next we extend the model to include "multilevel" effects. The multilevel model studies how the estimated coefficients vary across other characteristics of the municipalities. In this case, we consider how the effect

	Dependent variable:	
	Log Yields	Harvested Hectares
	(1)	(2)
GDDs / 1000	0.475***	1,700.306*
	(0.109)	(976.997)
KDDs / 1000	−2.989**	−23,179.330***
	(1.423)	(8,681.404)
Avg. Min.	−0.183***	−290.009
	(0.0183)	(335.665)
Precip. (m)	0.441***	−1,168.520*
	(0.076)	(677.845)
Precip.^{2} (m)	−0.494***	1,978.722**
	(0.099)	(854.580)
Observations	3,181	3,181
R^2	0.320	0.485
Adjusted R^2	0.290	0.462
Residual Std. Error (df = 3043)	0.364	14,412.800
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 2: Estimates for statistical models relating growing degree-days, killing degree-days, average minimum temperature, and precipitation to the logarithm of yields, and to harvested area, for the top 118 municipalities by production density. Stars (***) represent statistical significance levels, showing that most coefficients appear to have a relationship with production outputs.

of GDDs, KDDs, and average minimum temperature vary with elevation. Elevation is both an important determinant of coffee quality, and is a proxy for the variety of coffee grown: Brazil grows both Arabica and Robusta coffees, but does not report their production separately (until recent years).

The multilevel relationship is that:

$$\begin{aligned}
\log y_{it} &= \alpha_i + \gamma_i g_{it} + \kappa_i k_{it} + \mu_i m_{it} + \pi_i p_{it} + \psi_i p_{it}^2 + \epsilon_{it} \\
\gamma_i &= \gamma_0 + \beta_\gamma Elevation_i + \eta_{\gamma,i} \\
\kappa_i &= \kappa_0 + \beta_\kappa Elevation_i + \eta_{\kappa,i} \\
\mu_i &= \mu_0 + \beta_\mu Elevation_i + \eta_{\mu,i} \\
\pi_i &= \pi_0 + \beta_\pi Elevation_i + \eta_{\pi,i} \\
\psi_i &= \psi_0 + \beta_\psi Elevation_i + \eta_{\psi,i}
\end{aligned}$$

where the top line is the normal regression relationship, but with separate coefficients for each municipality i. The remaining lines relates all municipality coefficients together according to their varying elevations. The results are shown in table 3 and in a graphical form in figure 5.

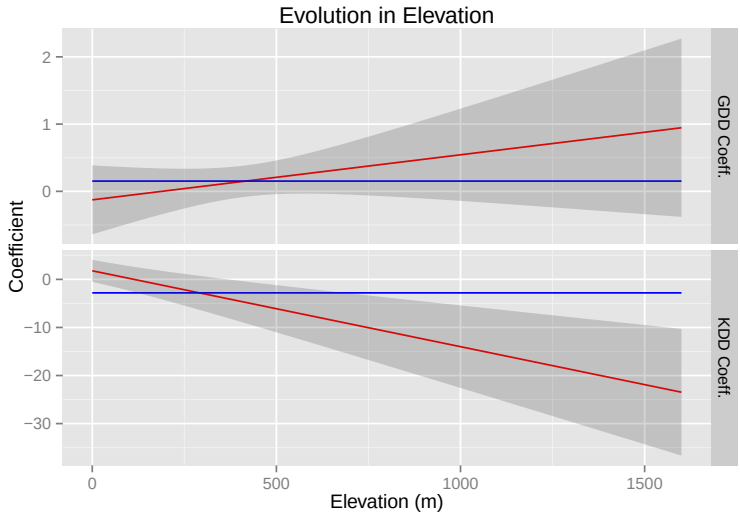


Figure 5: The effect of an additional GDD and KDD as these vary by elevation. As elevation increases, plants become more sensitive to temperatures. The effect of GDDs increases, though very slightly. The harmful effects of KDDs increase quickly.

0.2.6 Yield estimates under a warmer climate

We can apply the production model to weather produced from climate change. As a proxy for climate change, we estimate yields using historical weather data increased by 2°C. Precipitation values are left unchanged, since they show an unclear trend. This change produces several effects: it increases the number of GDDs benefiting yields, increases the number of KDDs harming yields, and increases average minimum temperature. The resulting balance between these three impacts is not evident *a priori*. The

	Dependent variable:	
	Log Yields	Harvested Hectares
	(1)	(2)
GDDs / 1000	0.208***	40.303
	(0.051)	(130.508)
Elev. GDDs / 1000	0.001***	2.110***
	(0.0002)	(0.657)
KDDs / 1000	−6.106***	−4,600.562***
	(0.516)	(725.931)
Elev. KDDs / 1000	−0.016***	−17.054***
	(0.002)	(3.653)
Avg. Min.	−0.183***	−25.750
	(0.018)	(34.334)
Elev. Avg. Min.	−0.00000**	−0.183
	(0.00000)	(0.183)
Precip. (m)	0.358***	−32.650
	(0.030)	(76.846)
Elev. Precip. (m)	0.0001	−0.164
	(0.0001)	(0.285)
Precip.^{2} (m)	−0.391***	−10.825
	(0.039)	(98.941)
Elev. Precip.^{2} (m)	0.0001	0.648*
	(0.0001)	(0.390)
Observations	42,141	42,161
R^2	0.378	0.651
Adjusted R^2	0.338	0.628
Residual Std. Error	0.538 (df = 39582)	4,282.486 (df = 39601)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 3: The effects of GDDs, KDDs, and average minimum, as each varies by elevation. While the estimates are not significant, they suggest increasing sensitivity to temperature in the form of both GDDs and KDDs as elevation increases. All municipalities in Brazil used.

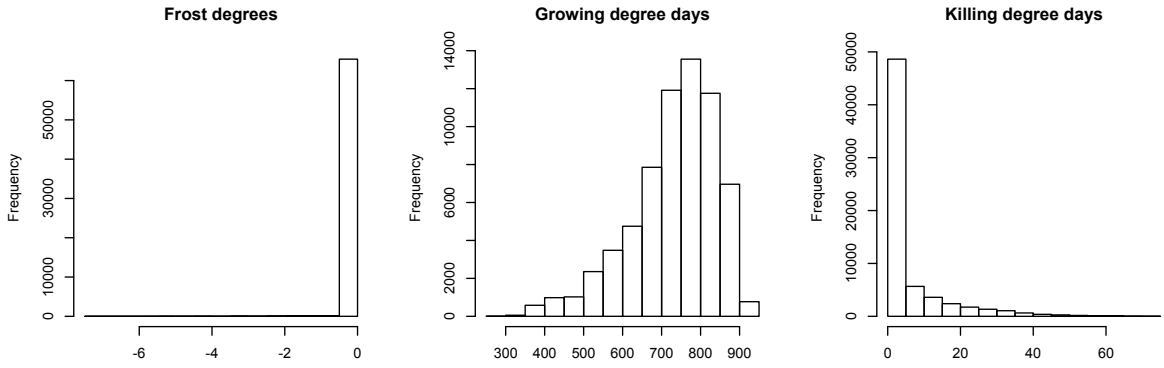


Figure 6: Growing degree day histograms, after an increase of 2°C.

figure below shows the distribution for municipality yields across Brazil, from observed data, and under climate changed weather predictions.

As shown in figure 2.6, the observed yields show wide variation. The blue distribution is shifted to the left, eliminating some of the most spectacular yields and lowering the average yield. The average yield in the warmer experiment is about 80% of the original yields (see figure 7).

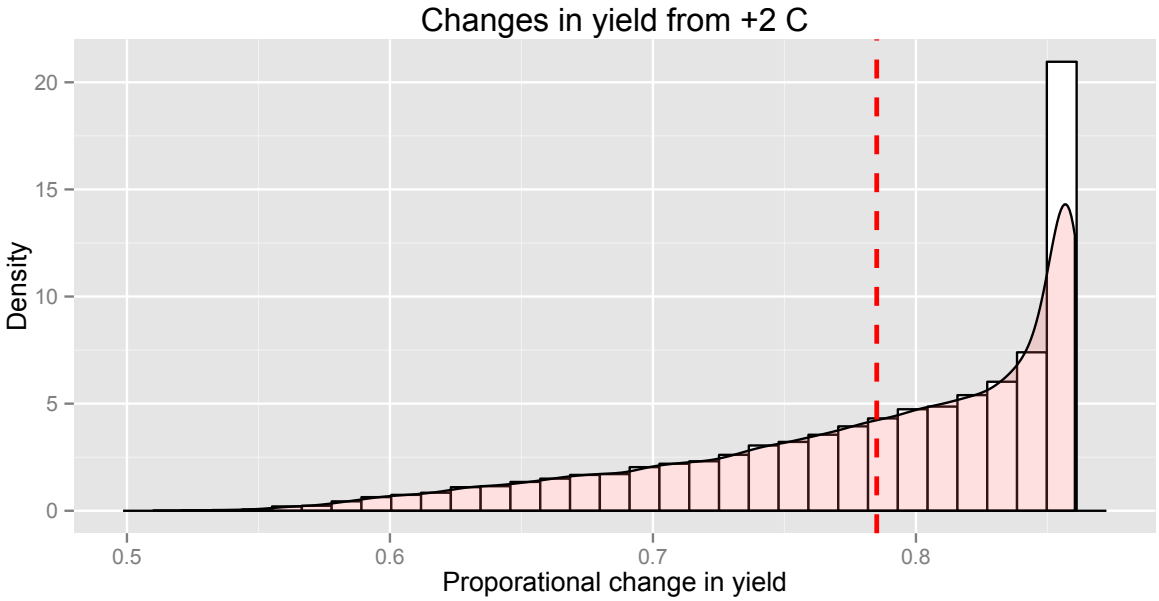


Figure 7: Distribution of the proportional change in yields, with a mean yield 79% of historical yields.

0.3 Global production

In this section, we estimate the a model like the one for Brazil for all countries. Using the intra-year production estimates in the coffee database, we estimate the relationship between country yields and weather. We use the temperature span of 0°C to 33°C for growing-degree days, as estimated for Brazil.

The first estimate is exactly analogous to the Brazil estimate, in that a single coefficient is estimated across all countries for the global average effect of GDDs, KDDs, frost degrees, and quadratic precipitation. This is reported in table 4 and shown schematically in figure 8.

0.3.1 Hierarchical model framework

It is reasonable to expect different countries to have different effects from temperatures. We could estimate each country independently, and this would be an "unpooled" model. However, we also want the model for one country to inform, to an extent supported by the data, the model for another country. To capture this, we will construct a "hierarchical model", where each country's sensitivity to temperature will be drawn from a common distribution, simultaneously estimating each country's parameters and the distribution across all of them.

Furthermore, we allow varieties in different regions to operate differently, as supported by the data. For example, where plentiful data supports a higher optimal growing temperature for Robusta, the model should represent this. If very little data is available, the predicted response should by default conform to an average for that region and variety. Finally, we want to incorporate higher resolution data where it is available. The municipality data in Brazil informs the same common parameters as the Brazil-specific

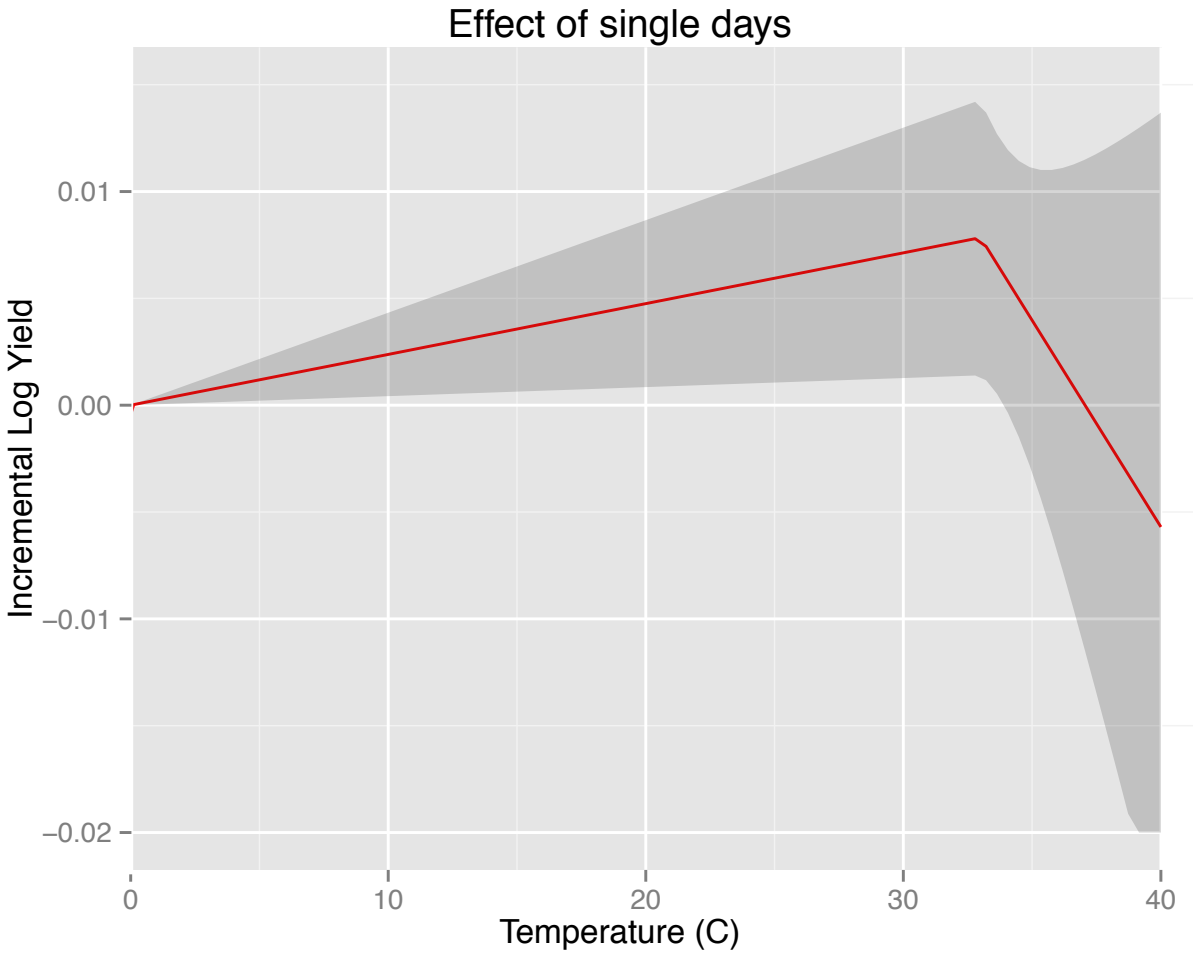


Figure 8: Pooled model growing degree-day plot.

	Log Yield	Production
GDD / 1000	0.238**	1,710.548
	(0.119)	(5,917.907)
KDD / 1000	−1.935	−3,955.098
	(1.786)	(43,378.870)
Frost Deg.	−0.005	284.772
	(0.008)	(1,550.480)
Year Precip	−3.454	707,932.900
	(12.928)	(483,967.400)
Year Precip^2	14.494	−10,991,768.000
	(135.355)	(6,955,772.000)
FE	*Region, variety*	*RegionVariety*
Trends	*Y*	*Y*
Errors	*Region*	*Region*
Observations	1,945	1,945
R^2	0.684	0.807
Adjusted R^2	0.676	0.802
Residual Std. Error (df = 1896)	0.441	33,325.380
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 4: Growing degree day model, pooled across all countries.

country-level yield data.

We have developed a technique for allowing this kind of data-driven multiple levels of aggregation and degrees of generalization, based on Bayesian Hierarchical Modeling (Gelman et al., 2014) and Inversion Theory (Menke, 2012). Under this technique, each country and sub-country region has its own parameters, but the parameters are further modeled as being related to each other. The hierarchical model is a direct extension of the statistical production model, which can be thought of as many different production models combined together.

Derivation of the hierarchical modeling system

Formally, we want to allow each variety in each country to have its own model, consisting of coefficients for growing degree-days, killing degree-days, average minimum temperature, and precipitation. The pooled model is as follows:

$$\log y_{it} = \alpha_i + \beta_v + \gamma g_{it} + \kappa k_{it} + \phi f_{it} + \pi p_{it} + \psi p_{it}^2 + \epsilon_{it}$$

while the partially pooled model starts with the unpooled relationship,

$$\log y_{ivt} = \alpha_i + \beta_v + \gamma_{iv} g_{it} + \kappa_{iv} k_{it} + \phi_{iv} f_{it} + \pi_{iv} p_{it} + \psi_{iv} p_{it}^2 + \epsilon_{ivt}$$

Consider the GDD coefficient for country i and variety v, γ_{iv}. To partial pool across countries for a given variety, this coefficient comes from a distribution of possible coefficient values, characterized by an unknown mean and standard deviation for that variety:

$$\gamma_{iv} \sim \mathcal{N}(\gamma_v, \tau_{\gamma_v})$$

Further, we partially pool these 'hyperparameters' as coming from a distribution across all varieties:

$$\gamma_v \sim \mathcal{N}(\gamma, \tau_\gamma)$$

We apply this for each parameter, $\gamma, \kappa, \phi, \pi, \psi$.

Estimating a partially-pooled model

Computationally, estimating this form of model can be very difficult. We construct an innovative framework for doing this using Ordinary Least-Squares matrix algebra.

The Gaussian relationships above, such as $\gamma_{iv} \sim \mathcal{N}(\gamma_v, \tau_{\gamma_v})$, are mathematically equivalent to the OLS-style relationship,

$$\gamma_{iv} = \gamma_v + \tau_{gamma_v}\eta \text{ with } \eta \sim \mathcal{N}(0,1)$$

Under OLS, error terms are members of a Gaussian distribution, $\epsilon_i \sim \mathcal{N}(0, \sigma_e^2)$. We represent the hyper-model for the γ coefficient with the OLS-style relationships

$$\begin{aligned}\gamma_{iv} &= \gamma_v + \epsilon_{iv}\\ \gamma_a &= \gamma_c + \epsilon_a\\ \gamma_r &= \gamma_c + \epsilon_r\end{aligned}$$

and similarly for the other coefficients. It is then possible to rewrite these and the original unpooled relationship to take the same form, with the same complete set of coefficients:

$\log y_{ivt}$	$= \alpha_i$	$+\gamma_{iv} g_{it}$		$+\cdots$
$\log y_{ivt}$	$= \sum_j \alpha_j \mathbf{1}_{j=i}$	$+ \sum_{ju} \gamma_{ju} g_{it} \mathbf{1}_{ju=iv}$	$+\gamma_a 0 + \gamma_r 0 + \gamma_c 0$	$+\cdots$
0	$= \sum_j \alpha_j 0$	$+ \sum_{ju} \gamma_{ju} \mathbf{1}_{ju=1a}$	$-\gamma_a 1 - \gamma_r 0 - \gamma_c 0$	$+\cdots$
0	$= \sum_j \alpha_j 0$	$+ \sum_{ju} \gamma_{ju} \mathbf{1}_{ju=2a}$	$-\gamma_a 1 - \gamma_r 0 - \gamma_c 0$	$+\cdots$
		$\vdots$		
0	$= \sum_j \alpha_j 0$	$+ \sum_{ju} \gamma_{ju} \mathbf{1}_{ju=1r}$	$-\gamma_a 0 - \gamma_r 1 - \gamma_c 0$	$+\cdots$
		$\vdots$		
0	$= \sum_j \alpha_j 0$	$+ \sum_j u\gamma_{ju} \mathbf{1}_{ju=1c}$	$-\gamma_a 0 - \gamma_r 0 - \gamma_c 1$	$+\cdots$
		$\vdots$		
0	$= \sum_j \alpha_j 0$	$+ \sum_j u\gamma_{ju} 0$	$+\gamma_a 1 + \gamma_r 0 - \gamma_c 1$	$+\cdots$
0	$= \sum_j \alpha_j 0$	$+ \sum_j u\gamma_{ju} 0$	$+\gamma_a 0 + \gamma_r 1 - \gamma_c 1$	$+\cdots$

The first line is the start of the original model to be estimated. The second line re-writes this with more systematically, and in such a way that "constant" terms can be set to zero for fictional observations. The remaining lines are fictional observations added to estimate the entire model.

We have built this approach into a tool for the R statistical package which is available at `https://github.com/eicoffee/hierlm`.

Figure 9 shows the effects of partial pooling at different levels. As the level of pooling increases, the

range of country-specific values is brought closer together.

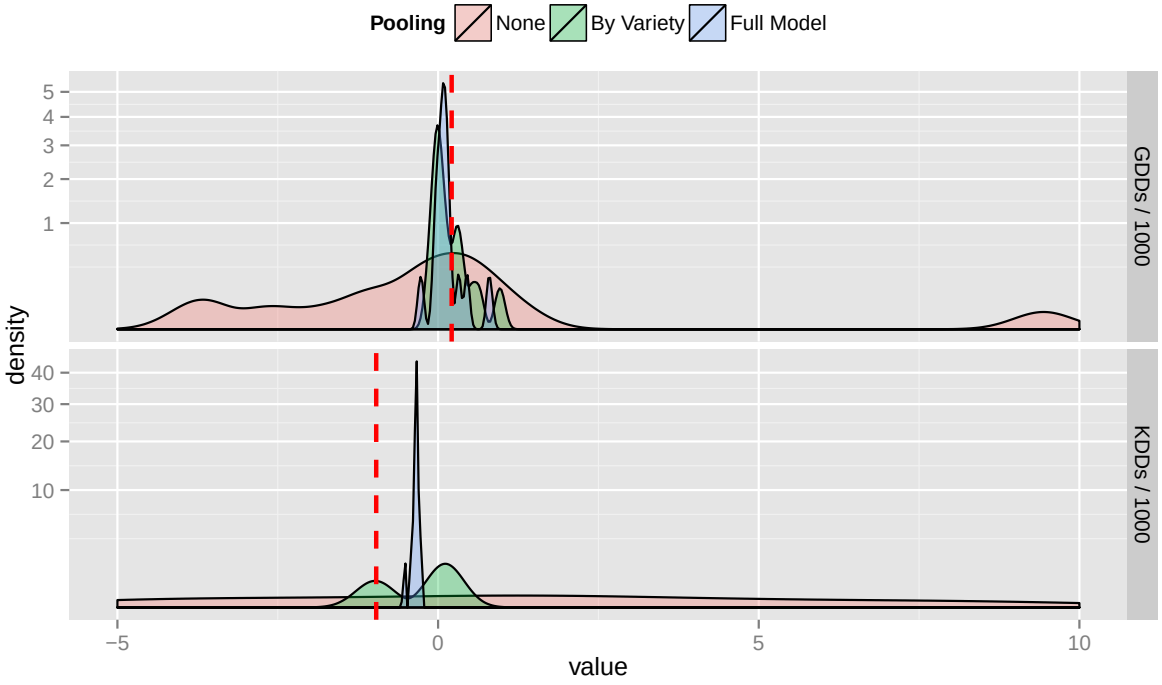


Figure 9: Distribution across countries of values for the GDD and KDD coefficients for different levels of pooling.

The results are shown in table 5. Only the hyperparameter means are shown. Each statistically significant country coefficient is listed in Appendix 0.5.5, and the remainder are in an online table at `http://eicoffee.net/`. The first column uses only observations at the country level. The second column places a prior on the Brazil coefficients, conforming to the Brazil municipality estimates above. These more-precise estimates then inform the global distribution for each coefficient, which in turn informs all of the countries, including Brazil.

0.3.2 Humidity

Humidity can have varying effects on coffee. The plant needs reasonably high levels of humidity during the flowering season to avoid floral atrophy, but humidity is also crucial to the development of coffee rust. For these reasons, the timing of high humidity levels appears to be particularly important. Here we see how Arabica coffee yields respond to a one-standard deviation increase in humidity during each particular month in the year leading up to harvest. Robusta appears to be less sensitive to humidity effects than Arabica.

Humidity data is from the NCEP CFSR. The reanalysis data is available at 1/12° resolution globally, which is then aggregated to the country-month level using weights from the coffee database. The values are reported as specific humidity at 6 hour intervals, which here is averaged over each month for the year prior to harvest.

	Dependent variable:	
	Countries only	
	(1)	(2)
GDDs / 1000 (Combined)	0.079	0.217**
	(0.123)	(0.095)
GDDs / 1000 (Arabica)	0.131	0.229**
	(0.112)	(0.103)
GDDs / 1000 (Robusta)	0.161	0.401***
	(0.152)	(0.133)
KDDs / 1000 (Combined)	−0.110	−1.801***
	(0.543)	(0.323)
KDDs / 1000 (Arabica)	−0.082	−1.731***
	(0.556)	(0.356)
KDDs / 1000 (Robusta)	−0.157	−1.766***
	(0.543)	(0.348)
Avg. Min. (Combined)	−0.077	−0.108
	(6.344)	(6.248)
Avg. Min. (Arabica)	−0.134	−0.152
	(7.147)	(7.164)
Avg. Min. (Robusta)	−0.114	−0.163
	(8.964)	(8.985)
Precip. (Combined)	−4.285	−2.124
	(5.792)	(2.390)
Precip. (Arabica)	−1.689	−0.156
	(6.058)	(3.254)
Precip. (Robusta)	−1.565	−0.279
	(5.971)	(3.403)
Precip.^{2}(Combined)	5.340	−5.530
	(82.317)	(28.605)
Precip.^{2}(Arabica)	21.749	11.218
	(79.174)	(37.825)
Precip.^{2}(Robusta)	12.794	0.264
	(88.198)	(42.271)
Observations	3,011	3,016
R^2	0.902	0.903
Adjusted R^2	0.885	0.886
Residual Std. Error	0.335 (df = 2561)	0.336 (df = 2566)
F Statistic	52.575*** (df = 450; 2561)	52.962*** (df = 450; 2566)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 5: Hierarchical model results, for the mean of the global distribution of coefficients for each parameter and each variety.

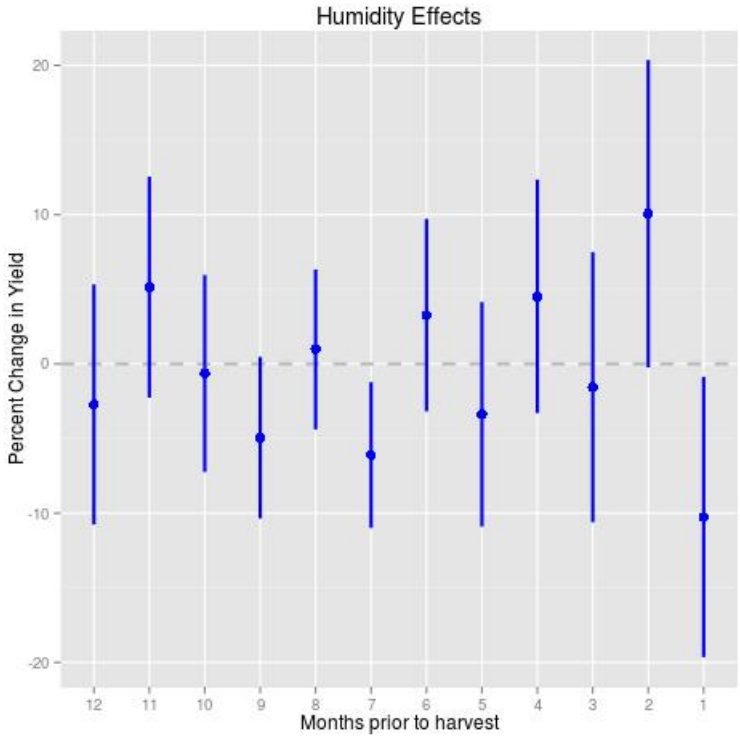


Figure 10: Arabica humidity effects. Only the humidity one and seven months before harvest are significant at 95% confidence.

Monthly effects of humidity are shown in figure 10, and the table of coefficients is in Appendix 0.5.3. The coefficients result from the following model:

$$log(y) = f(T) + \sum_{m=1}^{12} \beta_m q_m + h_c(t) + \alpha_c + \gamma_t + \epsilon_{ct}$$

where $f(T)$ is a non-linear function of temperature, estimated using the number of days spent in 1-degree C temperature bins, $h_c(t)$ is a country-specific linear time trend, α_c and γ_t are country and year fixed-effects. Each β_m is the effect of specific humidity m months prior to the beginning of harvest on log yield.

0.3.3 Interpreting empirical model results

Climate change impacts coffee production through many different channels. Foremost, climate change reflects changes in temperature and patterns of precipitation– that is, changes in climate mean changes in weather. The models above estimate the relationship between changes in weather and changes in yields, and then extrapolate those changes to their responses under climate change.

There are important differences between unexpected weather shocks and prolonged climate changes. Coffee farming will find ways to adapt to repeated shocks of higher temperatures, and we hope our estimates provide an upper bound on the production impacts of climate change. However, the evidence for such adaptation is limited. Burke and Emerick (2012) study maize in the United States, and while there is a clear potential for adaptation to warmer temperatures, they find almost no evidence of it. The reasons for this empirical result are unclear.

The effects that we measure of temperature on yields cannot be unambiguously interpreted as the biological response to temperatures. Temperatures could be simultaneously affecting other species that then affect coffee. For example, the harmful affects of average minimum temperature could reflect a greater capacity for coffee rust or the coffee berry borer to proliferate in these warmer years. It could also reflect decreased activity on the part of farmers on hot days.

Our results should be taken as representing a holistic effect as it has occurred in the past. The extent to which it will occur in the future may be up to us.

0.4 Pest modeling results

0.4.1 A rust model

We make a number of simplifications to study the fungus outbreaks. First, we will only consider the fungus's interaction with the host plant, even though it has been found to utilize other plants for different stages of its growth cycle. Secondly, we assume that the only factors influencing it spread are temperature and the health of the host plant, thereby ignoring wind and rain impacts that are also known to be important (Ferreira and Boley, 1991). Similarly ignored are higher order effects from the application of fungicide, where fungicide can also impact some flora and fauna that regulate the fungus, leading to potentially unpredictable disruptions in the natural system.

The model is initialized as a two-dimensional grid of farm space, each grid cell having a certain probability of an appearance of a fungal outbreak. For each time-step, chosen to be one month after examining the reproductive cycle of coffee rust, the outbreak will begin to increase in size as a function of both its current size, the temperature, the amount of host plants available and spread to neighboring grid-cells.

The temperature used in the model was obtained from surface temperature Reanalysis Data from the National Center of Atmospheric Research, spatially averaged over the area of Guatemala (without ocean cells) and temporally averaged to each month (NCAR, 2015). A time series of these average temperatures is shown in Figure 11, left. Here, one can note a relatively consistent seasonal amplitude of ˜2.5°C around a mean of ˜22°C, with a slight upwards skew.

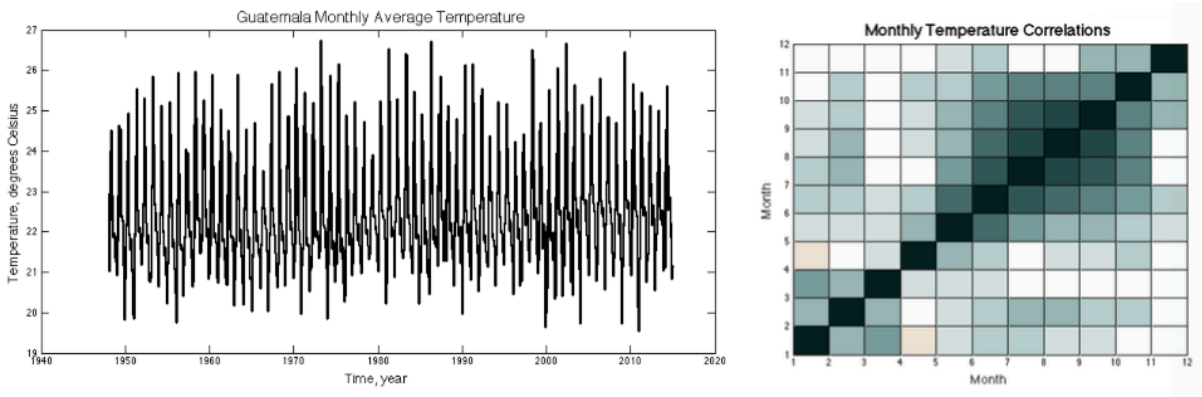


Figure 11: **Left:** Timeseries of temperatures. **Right:** Autocorrelation of monthly temperature.

The high level of correlation between months requires that the temperature selected in each monthly time step depend on the temperature in the previous month (see figure 11, right). In particular, the summer months are highly correlated, reaching correlations of 0.8 with the previous month in some cases. To account for this, the model is set up to draw a random season from the 67-year time series, employing a three month time series corresponding to an instance of summer, fall, winter or spring, depending on which is needed.

Our basic growth equation can be described by the following equation, with the basic assumption being that higher temperatures increase the growth rate of the fungus (at least at the temperatures seen in Guatemala).

$$N_{t+1} = N_t e^{r_i T_t / T_c}$$

where N is the population of a particular grid cell, r_i is the initial growth rate, T_t is the temperature at that specific time, T_c is the average temperature (over all months). Normally, the quantification of the growth rate is usually conducted with a consideration of both the daily maximum and daily minimum temperature (Magrath, 2014). However this was simplified for inclusion in our model.

When an additional population of fungus is created in subsequent time steps, it is distributed among the original and nearby grid-cells proportional to the health of the host plant in the new grid cell, the population of the source grid-cell and a multiplicative term similar to the prior growth equation. The maximum fungus population for each grid cell is 1, representing 100% infection of the host plant.

The disturbance of each grid-cell is also be subject to density-dependent pressure from predators, in this case the farmer spraying fungicide. Once a particular grid cell reaches a certain percentage of infection it is detected by the farmer. Detected, the population is decreased by a certain fraction, through the application of the fungicide. In addition, the fungus population will also decrease at a rate proportional to its current population and the relative health of the host species, independent of temperature.

$$N_{t+1} = N_t - N_t(F - 1)$$

where F is the percentage of available host plants for the fungus to grow on.

0.4.2 Experiments and results

To evaluate how temperature affects the spread of pests, experiments with three temperature scenarios were run. The first one uses each time step temperature (month) from the historical seasonal data for the region. The other two temperature scenarios considered global warming, one where the mean temperature was increased by 2°C and the other by 4°C. Those three temperature scenarios were run under two pest control conditions, the first one without any kind of pest control and the second one with a farmer's control by using fungicide that eliminated a fraction of the fungus when detected.

0.4.3 Historical temperature data without pest control

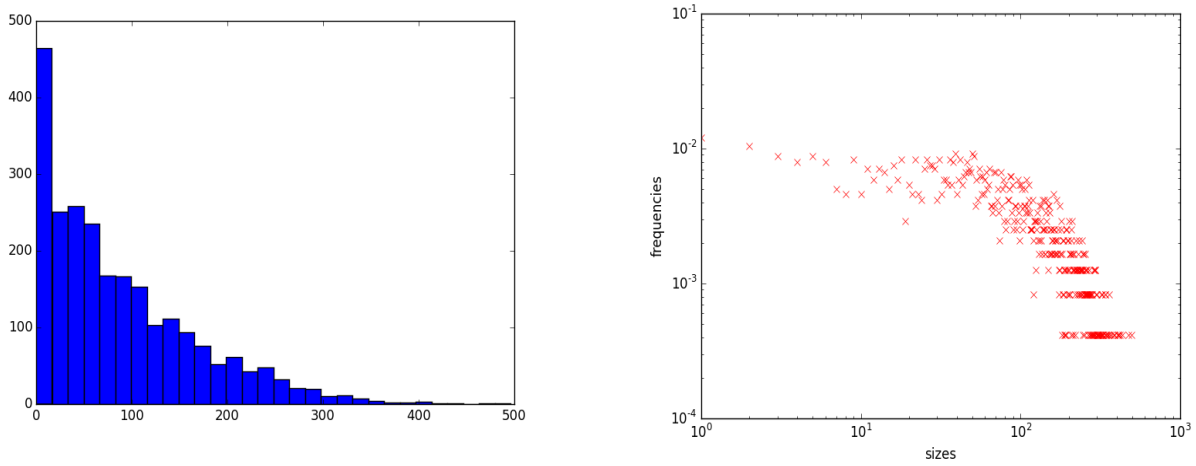


Figure 12: Historical temperature data without pest control. **Left:** Histogram of outbreak sizes. **Right:** Log-Log Plot of outbreak sizes.

We set the 'infected threshold' to about 0.3, from the simple fact that we found several instances in which leaves about 1/3 covered in coffee rust where considered 'heavily infected'. In figure 12 we have plotted the distribution of outbreak sizes, counting each heavily infected grid-cell at each time step. Please note that as we utilized a 30 by 30 grid cell, complete infection can be represented by a score of 900. Therefore, 490, the largest event, signifies that 54% of the crop is heavily infected.

The Log-Log result, though resembling a power law from 100 on, demonstrates some inconsistent behavior in smaller events. Indeed, in the histogram shelves can be observed where lower values have relatively the same probability. With no warming, the most common event is between 0 – 17, or between 0 and 2% of the crop (though this refers to a single month, not a harvest cycle). The largest event, at 54% of the crop is actually lower than the actual 70% loss in Guatemala in 2012, though there was not an indication of how this figure was calculated. Nevertheless, the model predicts scenarios where the fungus infects over 30% (300 sites) of the field.

Figure 13 shows the time series of the previous figure in months. It is clear that the system can be entrenched within certain domain (i.e., large or small events) for many years. As the behavior between 1750 and 2000, a 21-year period, consistently shows some of the highest outbreaks, while still containing intervening low events. This is possibly because the growth of fungus and plant, which are both tied to temperature, sometimes became more synchronous. However, this cannot be determined outright, and

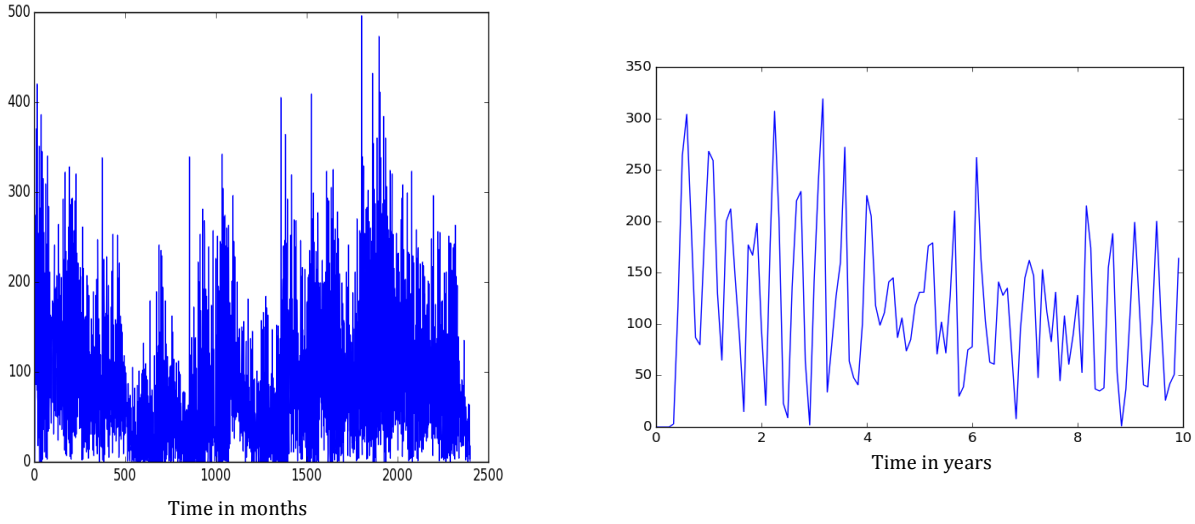


Figure 13: **Left:** Time series of outbreak sizes by month. **Right:** Time series of outbreak sizes by year.

thus for future study we might want to run for more time steps, to understand more fully the nature of this large- scale periodicity.

On smaller scales, it can also be noted that many of the largest events come directly after a period of relative calm, as the host plant has had a chance to regain health and provide much more nourishment to the attacking fungus. This small scale rebound, can be seen with more detail. Though the rapid up and down movement can be shown on a scale of a few months, there always seems to be a larger periodicity on the scale of a few years; however, the randomness in the system makes it difficult to conclude anything concrete.

0.4.4 Historical temperature data with pest control

Next, we implemented the farmer control, where obvious pest presence would immediately be sprayed with a fungicide and reduced to a fraction of its value in the next month. This fungicide and the necessary training to use it correctly may currently be absent within the poorer farms in the area. Thus, this allows us to see how implementation might change the situation in the future. In Figure 14, the distribution has a much smaller mean and median than the without the pest control measures. We note a stronger power law relationship, though like the last iteration, it is slightly concave down. This suggests that moderate events are marginally more likely than they would have been. However, the histogram might be slightly skewed by a larger prevalence of zero events. Additionally, one can note a larger spread than the previous model run for lower probability events.

Unexpectedly, the largest event, 643, signifying about 77% of the crop, is much higher than the previous iteration. This suggests that though fungicide keeps the fungus levels low for the average month, the healthy status of the host plant will make it so the correct temperature conditions or perturbation can cause a huge event, even before the farmer can react (here at a 1 month lag). This is very reminiscent of real world pest control experiences, where application can have unforeseen consequences, such as diminishing the population of a pest predator, and thus upsetting the natural structure of the system and allowing a pest to flourish later (Modern Farmer, 2014). However, catastrophic losses at a few points do not offset the considerable gains shown across the histogram.

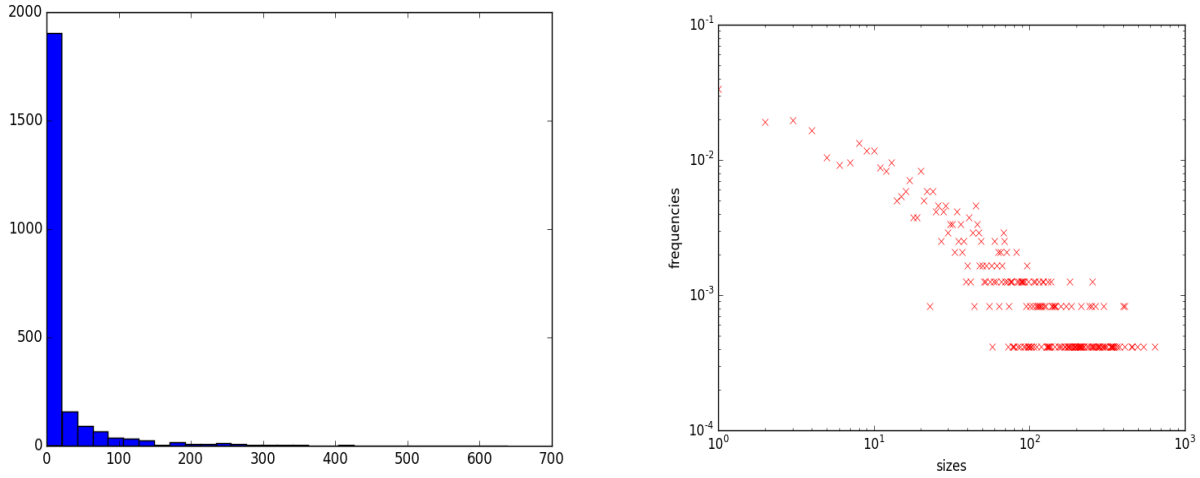


Figure 14: Historical temperature data with pest control. **Left:** Histogram of outbreak sizes. **Right:** Log-Log Plot of outbreak sizes.

Nevertheless, this line of thinking is further corroborated by figure 15, where decent periods of little activity are punctuated by huge events, a common feature in nonlinear spatial systems. However, when zoomed in to a period of 10 years, one can note the similarity between the control and non- control scenarios, where the lower bound in the control situation (within inter-month cycles) is replaced with 0.

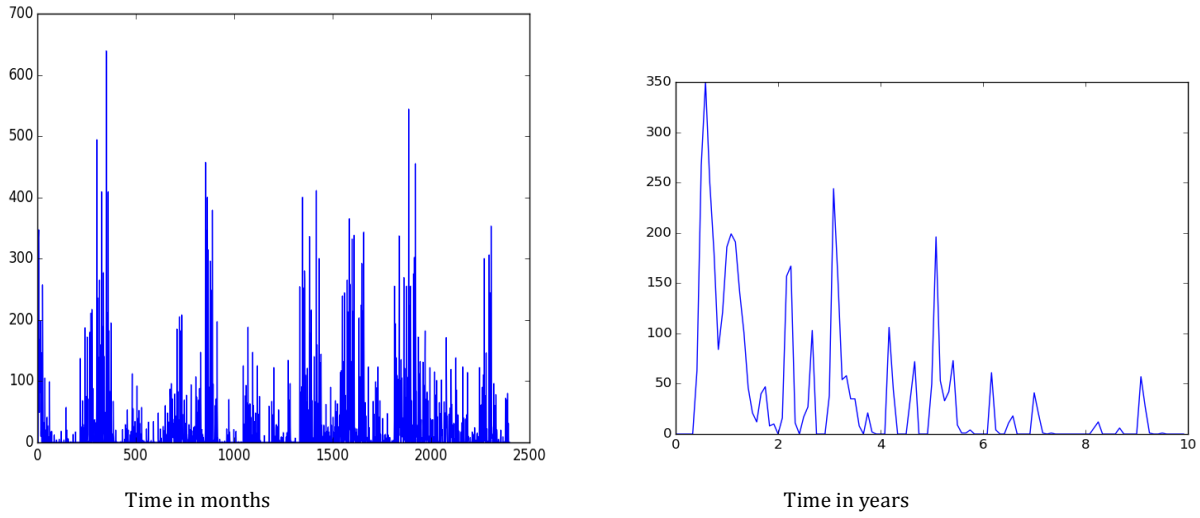


Figure 15: **Left:** Time series of outbreak sizes by month. **Right:** Time series of outbreak sizes by year.

0.4.5 2°C global warming temperature data without pest control

For the next run, we linearly increased the temperature of each month by 2 degrees, in order to represent possible regional warming over the next century. Increasing the temperatures to above normal, and thus often increasing the ability of the fungus to reproduce, causes the histogram of outbreak sizes to shift rightward. The log-log plot, while showing linear behavior for the right tail of the distribution, mimics this change. In an average month 10% of the crop is considered heavily infected, with the tail hitting about 75% of the crop as a maximum value. The time series (Figure 17) shows very few events with absolutely no fungus though the behavior in terms of both large-scale and small-scale periodicities does not seem to be drastically different.

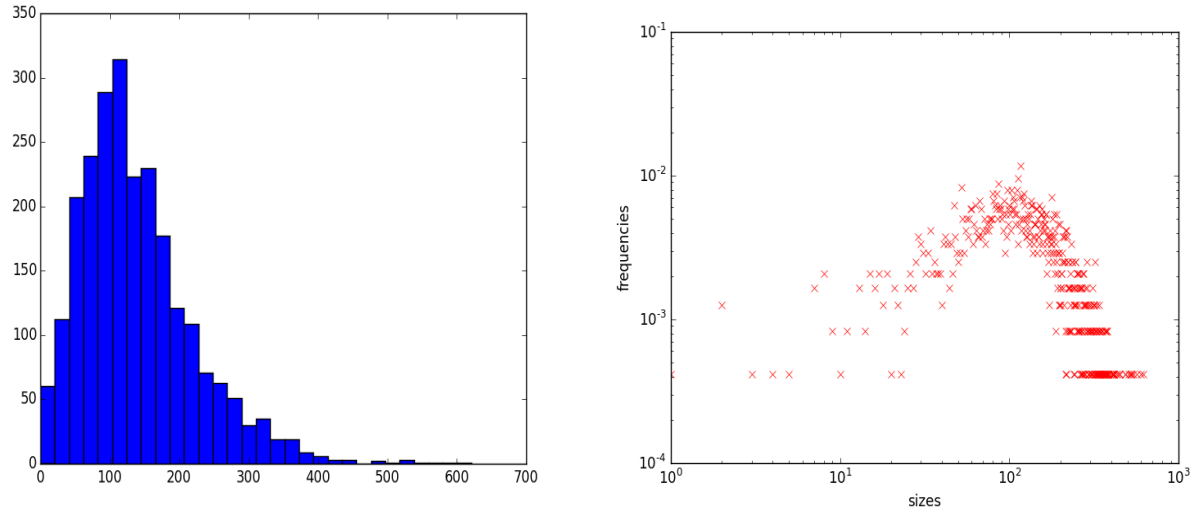


Figure 16: 2°C global warming temperature data without pest control. **Left:** Histogram of outbreak sizes. **Right:** Log-Log Plot of outbreak sizes.

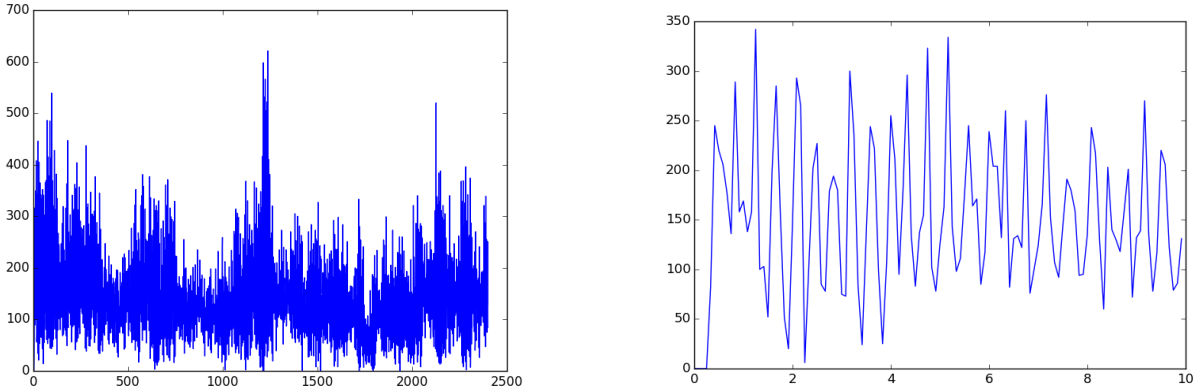


Figure 17: **Left:** Time series of outbreak sizes by month. **Right:** Time series of outbreak sizes by year.

0.4.6 2°C global warming temperature data with pest control

The addition of pest controls to the warmed scenario has a similar effect as we have noted in the previous iteration. The distribution begins to resemble a power law, however here with a slightly thinner tail. Nevertheless, even in the warmed environment the measures do a reasonable job of controlling the pests, with levels far below the untreated, cooler scenario. The Log-log plot, is however slightly more concave than the previous scenario with pest control.

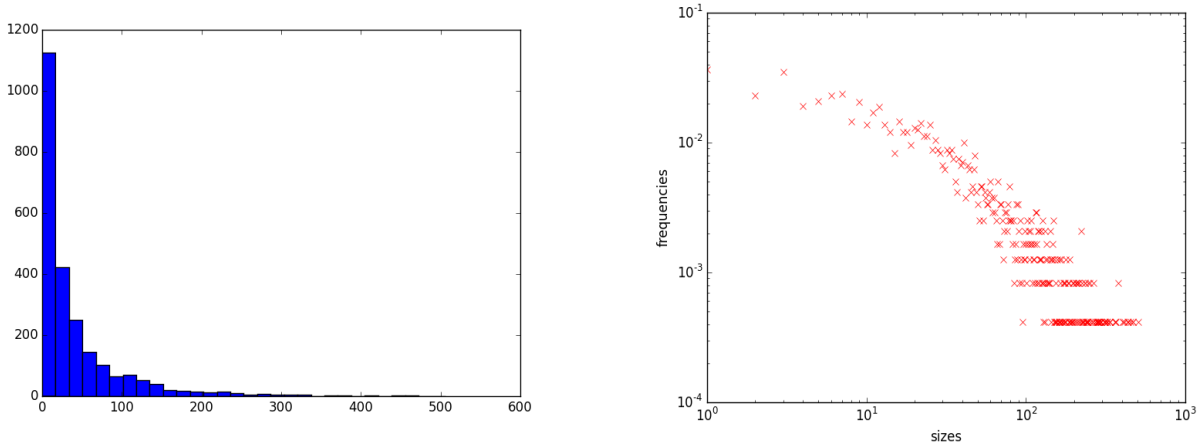


Figure 18: 2°C global warming temperature data with pest control. **Left:** Histogram of outbreak sizes. **Right:** Log-Log Plot of outbreak sizes.

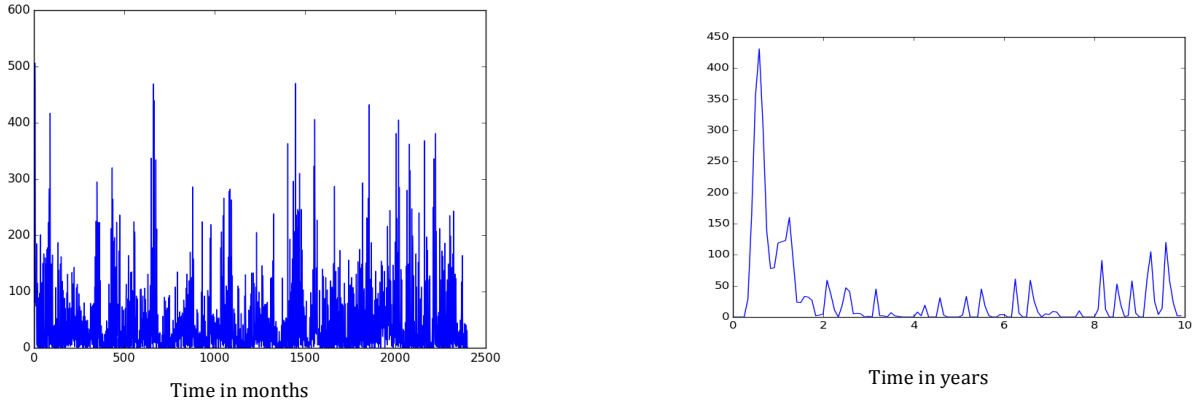


Figure 19: **Left:** Time series of outbreak sizes by month. **Right:** Time series of outbreak sizes by year.

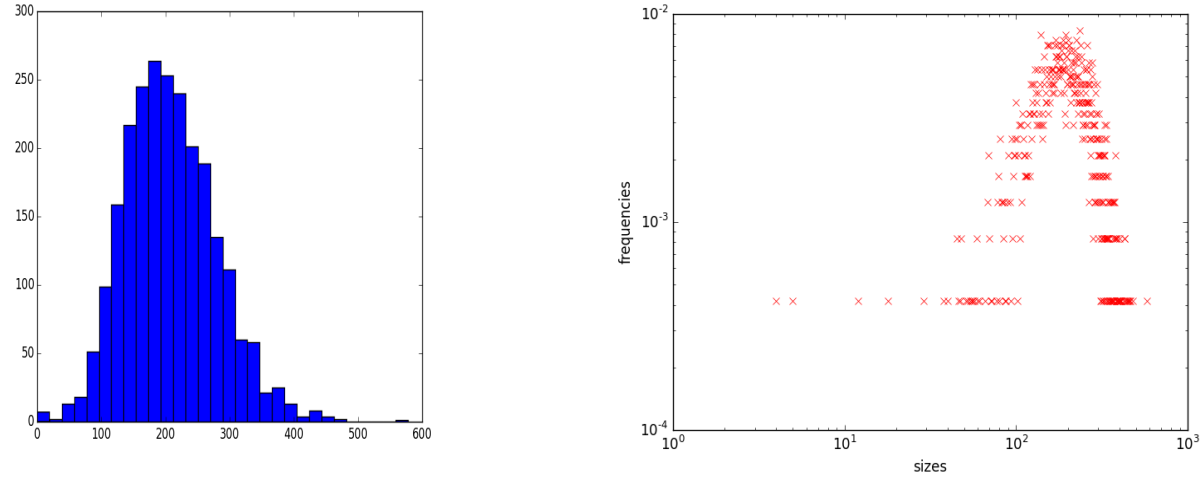
Results for a warming of 4°C are included in Appendix 0.4.7. Under these conditions and without pesticide, the health of the crop is so poor that it cannot maintain a full outbreak. Even with pesticide, it is impossible to full contain the disease.

0.4.7 Additional pest control results

4°C global warming temperature data without pest control

With 4°C of warming, the distribution becomes almost completely normal, continuing to move right, though with a smaller right tail. Even with a huge amount of warming, there were no instances above 600 or 66% of the crop. This suggests that with the average monthly infection around 20%, the plants are not healthy enough to sustain a super event like the size of one previously seen. This absolute limit does little to help the predictability on short time scales however.

Histogram of outbreak sizes and Log-log plot of outbreak sizes



Time series of outbreak sizes

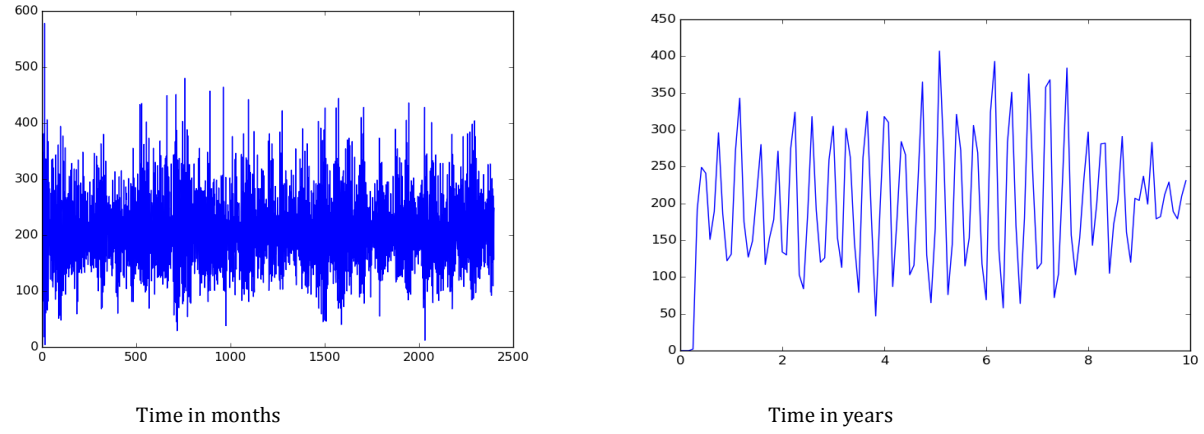


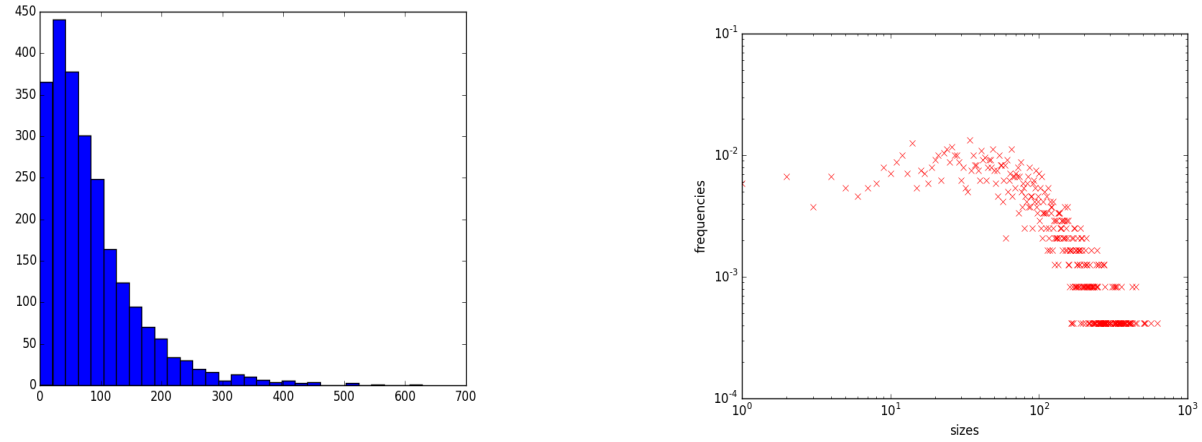
Figure 20: Results for 4°C warming without pest control.

4°C global warming temperature data with pest control

4°C of warming begins to offset the ability of fungicides to control the fungus population, the highest values in the histogram moves away from zero, shown by the concavity of the Log-log plot, as the curve continues to be thicker in areas than it was with less warming. The time series, as well as the other plots

itself begin to more resemble the model runs in uncontrolled but cooler environments, with the highest values still not going higher than 70%.

Histogram of outbreak sizes and Log-log plot of outbreak sizes



Time series of outbreak sizes

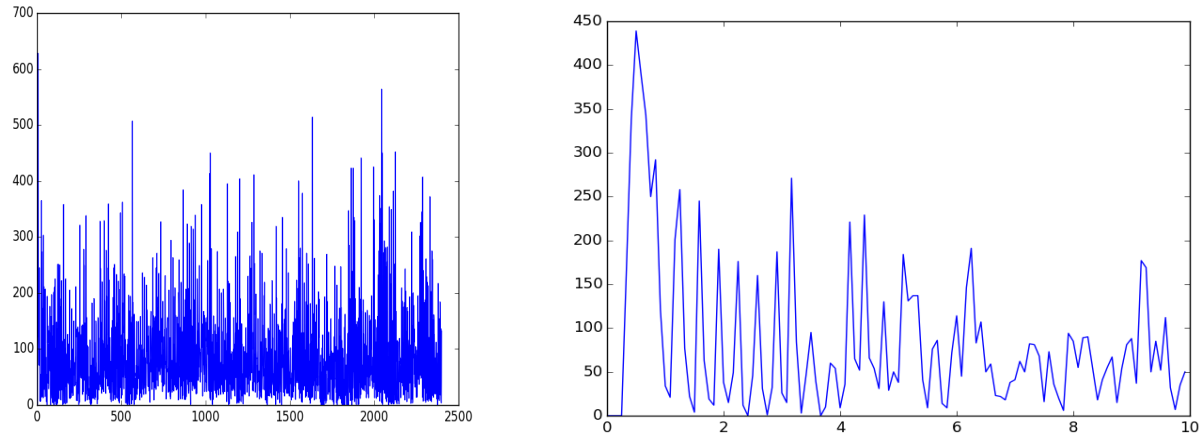


Figure 21: Results for 4°C warming with pest control.

0.4.8 Discussion

Throughout the drafting and modeling process, we made many other simplifications. We ignored rain and wind as possible spreading agents, instead opting for a random approach. We chose significant parameters such as our time step through very simplified observations of the fungus. We ignored the vegetation cycle (as Guatemala has a very defined wet and dry season which must affect plant growth), though this might be somewhat mitigated by the fact that we tied temperature to the growth of the host plant. We also made considerable simplifying assumptions about the qualities of fungicide application and fungus growth and spread.

Nevertheless, we believe that our results in the change of distribution are representative of what might occur in the real world, given a particular coffee field. We have noted that warming induces a rightward shift in event distribution. The subsequent health decline of the plants may inhibit huge shocks to the system, as the conditions are not ideal for a full fungus takeover. Pest control, while curtailing the

infection of an average month can lead to thicker tails and larger rare events. This is possibly because plants are kept at a healthier level, an ideal condition for a quick fungus take over and a drawback of an artificially controlled environment. Additionally, pest control appears to be efficient at compensating for the increased fungus growth rates caused by warming, as even in the 4°C warmer environment, it is able to bring the distribution back to the less disastrous approximate power law, albeit with a mode higher than zero. In the future, under the extreme scenario, it is very possible that some sort of artificial control will be necessary to continue to grow coffee in this region. This will possibly bring more complications and unpredictable dynamics that we cannot comment on with such a simple model.

0.5 Extra production analysis

0.5.1 Inputs to the ENSO analysis

Our projection is based on the most recent consensus projection of the NINO 3.4 index of ENSO, from International Research Institute for Climate and Society (2015). We try to apply reasonable values to the other indices, using the negative of NINO 3.4 for SOI, given its -0.6 correlation with NINO 3.4; a zero value for NAO, given its rapid shifts; and constant extrapolations for PDO and AMO at their most recent value, given the slow shifts in these signals.

Projecting these signals onto the principal component axes gives loadings of 1.3, 3.3, and 1.9, for the three components respectively.

To account for any spurious effect of our decision-making process, we estimate the values for 2014-15 as well, and report the difference.

0.5.2 Selecting temperature limits

0.5.3 Humidity

0.5.4 Harvest month effects

Figure 22 shows the estimated "effect" of harvesting in a given month on yields, from 1962 to 2011, after accounting for country-specific and monthly effects. The gradual increase reflects improvements in coffee production technology, but this increase is not without large shocks. An increase in yields between 1985 and 1990 was followed by a decrease and then another period of increased yields. Countries that harvest in different months also show different fortunes, with the greatest yields to countries that harvest in January and the lowest to those that harvest in February. Since the only country that harvests in January but not February is Colombia, this probably reflects the difference between Colombia yields and yields in other February-harvesting countries.

0.5.5 Hierarchical model coefficients

Only statistically-significant coefficients are listed below. The remaining are available online at `http://eicoffee.net/`.

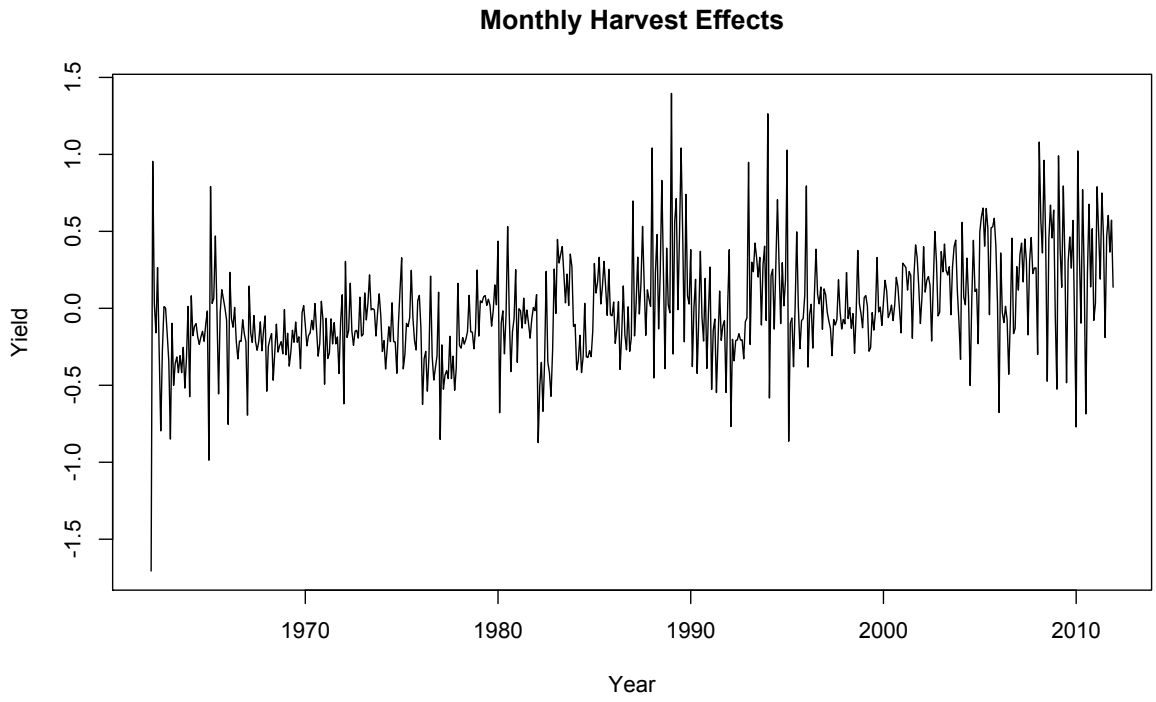


Figure 22: Monthly harvesting effects. Each point on this curve represents the difference in yields predicted by harvesting in a given month, according to coffee harvest calendars, after accounting for country-specific and month effects. Uses calendars from `https://www.sweetmarias.com/coffee.prod.timetable.php`

Low \High	28	29	30	31	32	33	34
-4					87.4211	87.4289	87.3933
-3							
-2						87.4290	
-1							
0				87.2986	87.4213	87.4290	87.3934
1						87.4290	
2					87.4212	87.4289	87.3933
3						87.4288	
4				87.2983	87.4210	87.4286	
5			87.0758	87.2979	87.4206		
6				87.2978	87.4205	87.4281	
7			87.0755	87.2978	87.4204		
8				87.2981			
9			87.0749	87.2979	87.4199		
10	86.7490		87.0737	87.2975			
11			87.0729	87.2975	87.4182		
12	86.7398		87.0700	87.2954			
13							
14							
15	86.6988		87.0369	87.2645			

Table 6: F-statistics for a growing degree-day and killing degree-day model of coffee production, across all countries. The highest F-stats use a maximum temperature of 30°C and a minimum temperature between -3°C and 1 °C.

	Dependent variable:	
	Countries only	
	(1)	(2)
GDDs / 1000, Liberia (Robusta)	0.515**	0.743***
	(0.213)	(0.202)
GDDs / 1000, Gabon (Robusta)	0.223	0.448**
	(0.215)	(0.204)
GDDs / 1000, Yemen (Arabica)	0.274	0.368**
	(0.189)	(0.183)
GDDs / 1000, Benin (Robusta)	0.146	0.409**
	(0.221)	(0.207)
GDDs / 1000, Cuba (Arabica)	0.222	0.322*
	(0.194)	(0.189)
GDDs / 1000, Angola (Robusta)	0.121	0.354*
	(0.217)	(0.205)
GDDs / 1000, Malaysia (Robusta)	0.266	0.495**
	(0.220)	(0.209)
GDDs / 1000, Brazil (Combined)	0.079	0.158***
	(0.208)	(0.052)
GDDs / 1000, Guinea (Robusta)	0.356*	0.603***
	(0.199)	(0.185)
GDDs / 1000, Nigeria (Robusta)	0.377*	0.659***
	(0.212)	(0.197)
GDDs / 1000, Suriname (Combined)	0.346*	0.484**

	Dependent variable:	
	Countries only	
	(1)	(2)
	(0.204)	(0.189)
GDDs / 1000, Zambia (Arabica)	0.217	0.300*
	(0.178)	(0.173)
GDDs / 1000, Paraguay (Arabica)	0.248	0.405***
	(0.165)	(0.156)
GDDs / 1000, Guyana (Robusta)	0.140	0.374*
	(0.223)	(0.211)
GDDs / 1000, Congo (Robusta)	0.145	0.382*
	(0.215)	(0.203)
KDDs / 1000, Cambodia (Combined)	−0.112	−1.798***
	(0.567)	(0.363)
KDDs / 1000, Ethiopia (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Cameroon (Combined)	−0.111	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Ghana (Robusta)	−0.180	−1.787***
	(0.568)	(0.386)
KDDs / 1000, Saudi.Arabia (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Guatemala (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Guatemala (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Dominica (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Liberia (Robusta)	−0.123	−1.732***
	(0.568)	(0.386)
KDDs / 1000, Gabon (Robusta)	−0.155	−1.764***
	(0.568)	(0.386)
KDDs / 1000, Gabon (Combined)	−0.110	−1.800***
	(0.568)	(0.364)
KDDs / 1000, Yemen (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Yemen (Arabica)	−0.078	−1.728***
	(0.581)	(0.394)
KDDs / 1000, Jamaica (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Samoa (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Kenya (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Kenya (Combined)	−0.114	−1.804***
	(0.568)	(0.364)
KDDs / 1000, India (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Saint.Lucia (Combined)	−0.110	−1.801***

	Dependent variable:	
	Countries only	
	(1)	(2)
	(0.568)	(0.364)
KDDs / 1000, Rwanda (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Peru (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Vanuatu (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Malawi (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Benin (Robusta)	−0.156	−1.754***
	(0.565)	(0.384)
KDDs / 1000, Benin (Combined)	−0.114	−1.773***
	(0.559)	(0.358)
KDDs / 1000, Cuba (Arabica)	−0.076	−1.725***
	(0.581)	(0.394)
KDDs / 1000, Togo (Robusta)	−0.244	−1.827***
	(0.560)	(0.380)
KDDs / 1000, Tonga (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Indonesia (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Mauritius (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Angola (Combined)	−0.109	−1.799***
	(0.568)	(0.364)
KDDs / 1000, Angola (Robusta)	−0.159	−1.768***
	(0.568)	(0.386)
KDDs / 1000, Trinidad.and.Tobago (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Nicaragua (Arabica)	−0.084	−1.733***
	(0.581)	(0.394)
KDDs / 1000, Malaysia (Robusta)	−0.159	−1.768***
	(0.568)	(0.386)
KDDs / 1000, Mozambique (Combined)	−0.111	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Uganda (Combined)	−0.111	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Brazil (Combined)	−0.110	−1.971***
	(0.568)	(0.309)
KDDs / 1000, Guinea (Robusta)	−0.101	−1.703***
	(0.566)	(0.384)
KDDs / 1000, Panama (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Costa.Rica (Arabica)	−0.081	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Nigeria (Robusta)	−0.085	−1.674***

	Dependent variable:	
	Countries only	
	(1)	(2)
	(0.562)	(0.382)
KDDs / 1000, Ecuador (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, El.Salvador (Arabica)	−0.081	−1.729***
	(0.581)	(0.393)
KDDs / 1000, Puerto.Rico (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Thailand (Combined)	−0.109	−1.796***
	(0.567)	(0.363)
KDDs / 1000, Thailand (Robusta)	−0.165	−1.773***
	(0.568)	(0.386)
KDDs / 1000, Haiti (Arabica)	−0.087	−1.731***
	(0.580)	(0.393)
KDDs / 1000, Belize (Combined)	−0.110	−1.799***
	(0.568)	(0.364)
KDDs / 1000, Sierra.Leone (Robusta)	−0.239	−1.835***
	(0.563)	(0.383)
KDDs / 1000, Philippines (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Timor.Leste (Combined)	−0.109	−1.799***
	(0.568)	(0.364)
KDDs / 1000, Colombia (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Burundi (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Burundi (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Fiji (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Madagascar (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Nepal (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Suriname (Combined)	−0.089	−1.779***
	(0.568)	(0.364)
KDDs / 1000, Zambia (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Papua.New.Guinea (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Zimbabwe (Arabica)	−0.094	−1.742***
	(0.581)	(0.393)
KDDs / 1000, New.Caledonia (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, New.Caledonia (Arabica)	−0.082	−1.731***
	(0.581)	(0.394)
KDDs / 1000, Paraguay (Arabica)	−0.043	−1.660***

	Dependent variable:	
	Countries only	
	(1)	(2)
	(0.571)	(0.387)
KDDs / 1000, Guyana (Robusta)	−0.157	−1.766***
	(0.568)	(0.386)
KDDs / 1000, Guyana (Arabica)	−0.081	−1.730***
	(0.581)	(0.394)
KDDs / 1000, Guyana (Combined)	−0.111	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Honduras (Arabica)	−0.084	−1.733***
	(0.581)	(0.394)
KDDs / 1000, Myanmar (Combined)	−0.110	−1.800***
	(0.568)	(0.364)
KDDs / 1000, Mexico (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
KDDs / 1000, Congo (Robusta)	−0.165	−1.773***
	(0.568)	(0.386)
KDDs / 1000, Congo (Combined)	−0.111	−1.800***
	(0.568)	(0.364)
KDDs / 1000, Sri.Lanka (Combined)	−0.108	−1.793***
	(0.567)	(0.363)
KDDs / 1000, Comoros (Combined)	−0.110	−1.801***
	(0.568)	(0.364)
Avg. Min., Liberia (Robusta)	−0.817***	−0.873***
	(0.141)	(0.140)
Avg. Min., Gabon (Robusta)	−0.642***	−0.714***
	(0.181)	(0.180)
Avg. Min., Yemen (Combined)	0.402*	0.369*
	(0.208)	(0.207)
Avg. Min., Jamaica (Arabica)	0.297**	0.269**
	(0.117)	(0.116)
Avg. Min., Kenya (Arabica)	−1.325***	−1.369***
	(0.288)	(0.288)
Avg. Min., Kenya (Combined)	−0.755***	−0.794***
	(0.156)	(0.154)
Avg. Min., Malawi (Arabica)	−0.255*	−0.288**
	(0.141)	(0.140)
Avg. Min., Angola (Combined)	−0.367*	−0.400**
	(0.199)	(0.198)
Avg. Min., Angola (Robusta)	0.218**	0.178*
	(0.110)	(0.108)
Avg. Min., Malaysia (Robusta)	2.766***	2.680***
	(0.164)	(0.162)
Avg. Min., Brazil (Combined)	−0.077	−0.091***
	(34.091)	(0.020)
Avg. Min., Guinea (Robusta)	0.329***	0.300**
	(0.125)	(0.124)
Avg. Min., El.Salvador (Arabica)	−0.525***	−0.526***

	Dependent variable:	
	Countries only	
	(1)	(2)
	(0.145)	(0.145)
Avg. Min., Sierra.Leone (Robusta)	−1.196***	−1.197***
	(0.147)	(0.147)
Avg. Min., Suriname (Combined)	−1.532***	−1.559***
	(0.174)	(0.172)
Avg. Min., Zambia (Arabica)	−0.204	−0.237*
	(0.125)	(0.123)
Avg. Min., Congo (Robusta)	−1.389***	−1.441***
	(0.155)	(0.154)
Avg. Min., Sri.Lanka (Combined)	−0.426***	−0.402***
	(0.140)	(0.139)
Precip., Brazil (Combined)	−4.285	0.347***
	(6.691)	(0.030)
Precip., Suriname (Combined)	−12.378*	−10.271**
	(6.552)	(4.007)
Precip.^{2}, Brazil (Combined)	5.340	0.366***
	(88.871)	(0.039)
Observations	3,011	3,016
R^2	0.902	0.903
Adjusted R^2	0.885	0.886
Residual Std. Error	0.335 (df = 2561)	0.336 (df = 2566)
F Statistic	52.575*** (df = 450; 2561)	52.962*** (df = 450; 2566)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 7: Humidity Effects

	Dependent variable:
Month prior to harvest	log(yield)
1	−8.562
	(12.703)
2	17.386
	(12.509)
3	−2.607
	(13.406)
4	−23.317*
	(12.756)
5	4.781
	(13.223)
6	−30.035**
	(12.008)
7	15.021
	(14.797)
8	−14.813
	(16.496)
9	19.024
	(16.429)
10	−6.111
	(17.747)
11	35.636*
	(18.228)
12	−33.730**
	(15.444)
Observations	738
R^2	0.895
Adjusted R^2	0.881
Residual Std. Error	0.191 (df = 653)
F Statistic	66.164*** (df = 84; 653)
Note:	*p<0.1; **p<0.05; ***p<0.01

0.6 Suitability analysis details

Given any environmental condition, we can use Bayes rule to provide a empirical estimate of suitability. We write Bayes rule as an odds ratio:

$$\frac{p(\text{coffee} = 1|\vec{x})}{p(\text{coffee} = 1)} = \frac{p(\vec{x}|\text{coffee} = 1)}{p(\vec{x})}$$

The left-hand-side describes the ratio of the probability of coffee in a region given the observed conditions, to the probability of coffee generally. If this is greater than 1, the area is more suitable than the average location.

To calculate the coffee probability, the right-hand-side describes a ratio between the distribution of a property across harvested areas, and the distribution of that property across the entire region. As conditioning data, we use soil properties, climatic properties, elevation, and latitude.

Climatic and soil properties are not mutually independent, complicating our ability to calculate this ratio given the large number of properties we have available. We use the statistical "copulas" technique to disentangle the marginal distributions of each property from their dependence structure (Nelsen, 2013).

0.6.1 Comparing MaxEnt and Bayesian odds techniques

Both the MaxEnt and Bayesian techniques are sophisticated and represent the uncertainty of their result with high integrity. Table 9 provides a comparison of the main strengths and weaknesses of the two techniques.

	MaxEnt	**Bayesian Odds**
Form of observation data	A set of point locations, for where the species was observed	A weighted grid of presence across all space
Use of environmental data	Mean and other moments of the environmental data are used, based on the constraints chosen	The full distribution of values for any environmental indicators are included
Use of constraints	Constraints are central to the technique	Constraints are not used
Result uncertainty	Fully maintained	Fully maintained
Key simplification	Only chosen constraints describe the distribution of environmental indicators	Dependence between indicators is assumed to be captured by a rank correlation.

Table 9: Comparison between the features of MaxEnt and Bayesian odds methods.

MaxEnt has been used to study species suitability for a long time, and is designed for situations where a species is observed only at particular locations. The spatial coffee database gives us a much clearer picture of where coffee currently is grown and is not.

MaxEnt is most appropriate when there are underlying motivations for the constraints that are used as a central part of the method: for example, a common constraint is to require that that the mean and variance of temperatures for observed coffee match the mean and variance for future coffee. Unfortunately, MaxEnt constraints are often chosen arbitrarily and without a physical foundation. The Bayesian odds

approach incorporates the entire distribution of environmental indicator values where coffee is grown, rather than a small collection of moments.

0.6.2 Baseline Bayesian odds map

The result of the Bayesian odds procedure for current coffee suitability is shown in figure 23. Dark green regions (high suitability) are rare, unlike the analyses by GAEZ and Bunn et al.. While they typically match areas of actual coffee growth (in Brazil, Colombia, and Central America), there are several places where there are large mismatches (in North Africa and Western India). While this provides a high resolution and data-driven map, it cannot stand alone.

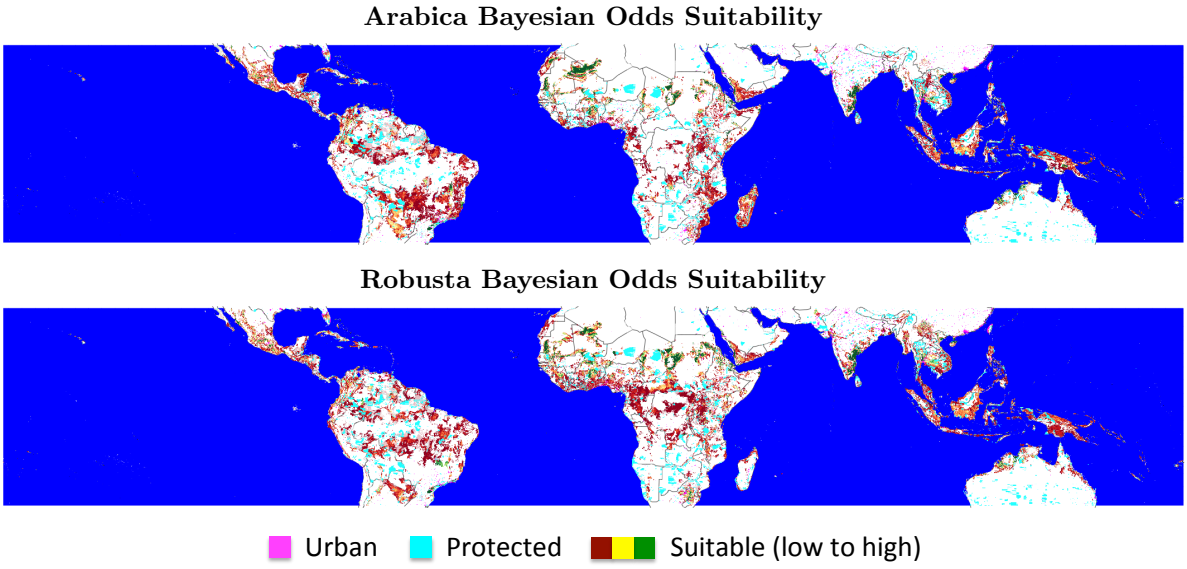


Figure 23: Suitability for Arabica coffee (top) and Robusta coffee (bottom). Colors range from red (slight suitability odds) to yellow to green (very strong suitability odds). The map also shows protected areas (cyan), urban areas (purple), and managed areas (faded).

0.6.3 Use of Copulas in the Bayesian odds measure

We use a Gaussian copula, which captures the correlation between the various properties.

To incorporate a new property, we determine its unweighted distribution across the entire region from 30°N to 30°S. Then we create a weighted distribution, with properties from the region weighted by harvested area. Finally, we calculate Spearman's rho, between the new property and all existing properties, to represent the dependence structure.[1]

Then, to determine $p(\vec{x})$ and $p(\vec{x}|\text{coffee} = 1)$ for a given location, we reverse the normal copula process. In this case, we determine the span in $\vec{u}$-space (rank space) that a small region of $\vec{x}$-space represents $(\vec{x} \pm \vec{\Delta x})$, using each marginal distribution and the probability integral transform. If there is very little

[1] Either Spearman's rho and Kendall's tau can be used in this process. We use Spearman's rho because it has a more linear relationship with the Gaussian copula's correlation matrix.

mass in the marginal distribution in the region of x_i, the corresponding Δu_i will be small. Then we evaluate

$$\int_{\vec{\Delta u}} c_R^{\text{Gauss}}$$

Above, c_R^{Gauss} is the Gaussian copula, which can be written as,

$$c_R^{\text{Gauss}}(u) = \frac{1}{\sqrt{\det R}} \exp\left(-\frac{1}{2}\begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots \\ \Phi^{-1}(u_d)\end{pmatrix}^T \cdot (R^{-1} - \mathbf{I}) \cdot \begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots \\ \Phi^{-1}(u_d)\end{pmatrix}\right)$$

where Φ^{-1} is the inverse cumulative distribution function of a standard normal (Arbenz, 2013), and R is the matrix of correlations, equal to $2\sin\rho_{ij}\frac{\pi}{6}$ for each Spearman's rho, ρ_{ij}, between property i and property j.[2]

0.6.4 Incorporating biological process

The Global Agro-ecological Zones (GAEZ) project uses biologically-motivated calculations to estimate suitability. GAEZ suitability indexes are normalized to be between 0 and 100, so a comparison between the Bayesian results and GAEZ requires constructing a common scale. We do this by comparing the results in ranks, rather than levels. In other words, we look for differences in the percentile quality of land (see figure 24).

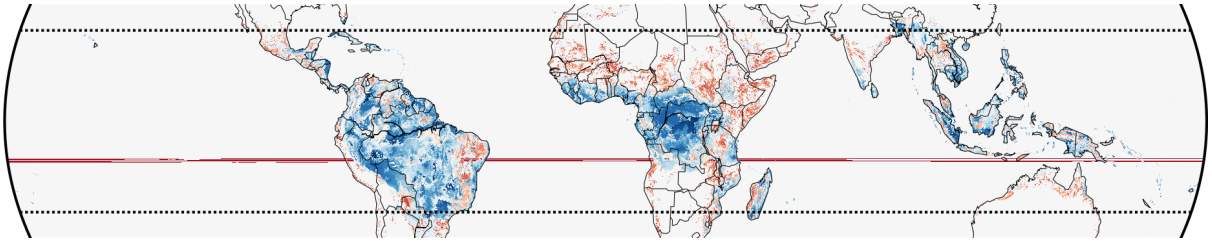


Figure 24: Comparison between GAEZ and the Bayesian odds technique for Arabica. Blue regions have greater quantile suitability in GAEZ than for the Bayesian odds approach; red regions show lower suitability in GAEZ, and white regions agree.

Some areas match closely (southern Brazil, Colombia, and parts of Indonesia), while GAEZ attributes suitability to large regions not supported by the Bayesian methods, such as Amazonian and Congo rainforest. This indicates a complementarity between GAEZ and the Bayesian odds approach, where GAEZ provides physical constraints while the Bayesian approach forces the results to match observed data.

Computing a combined metric

We combine the two approaches by mapping the following:

$$s(x,y) = p(x,y)\frac{b(x,y)}{1+b(x,y)}$$

[2]See `http://www.mathworks.com/help/stats/copulas-generate-correlated-samples.html#buqq6py`.

This attributes zero suitability where either approach specifies it, and otherwise allows them to reinforce each other. The results are shown in figure 5.7. It also normalizes the result to match GAEZ 0 - 100 scale.

The combined result shows high suitability in many fewer places, scattered based on where both techniques support their suitability. This provides a stronger basis for identifying the shifts in suitability, conservatively matched to only the highly suitable regions.

0.6.5 Suitability comparison with Bunn et al.

A recent paper by Bunn et al. (2015) uses data mining methods, such as MaxEnt, on coffee-growing presence at 42 000 individual farms to estimate suitability. Above, we build upon this work by incorporating the coffee presence map from their paper into our database. We also use the same collection of 19 bioclimactic variables, on top of which we add soil variables, and we extend the study of future uncertainty by using 12 additional global climate model results. While we believe that our approach, based on Bayesian updating of presence and absence information, is better grounded theoretically and less arbitrary than their MaxEnt and other data-mining techniques, Bunn et al. provides an important comparison for our results.

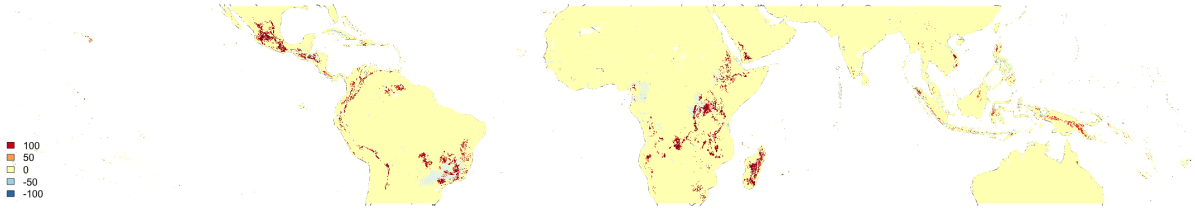


Figure 25: Comparison between Bunn et al. (2015) and the combined Bayesian-GAEZ approach. Blue areas have higher suitability in the baseline map produce by Bunn et al., while red is higher using our approach.

Figure 25 displays a comparison of current suitability between the two methods. Most of the world in this figure is colored yellow, where both techniques specify very little suitability. Some areas, such as Brazil and Kenya, show differing patterns between the two approaches. In these cases, our approach produces a result that more closely matches the patterns in the coffee database.

0.6.6 Changes in suitability by country for GAEZ

Country	Baseline (1000 Ha)	A2 2050 (1000 Ha)	Change (1000 Ha)	%
Angola	63738	40508	-23230	(-36%)
Argentina	9047	12173	+3126	(+35%)
Australia	13870	7593	-6277	(-45%)
Bahamas	3487	1821	-1666	(-48%)
Bangladesh	11677	298	-11379	(-97%)
Belize	2652	1642	-1010	(-38%)
Benin	4129	0	-4129	(-100%)
Bhutan	0	1216	+1216	(new)

Country	Baseline (1000 Ha)	A2 2050 (1000 Ha)	Change (1000 Ha)	%
Bolivia	76211	7791	-68420	(-90%)
Brazil	785103	235221	-549882	(-70%)
Cambodia	15559	1084	-14475	(-93%)
Cameroon	39370	34157	-5213	(-13%)
Central African Republic	56584	33494	-23090	(-41%)
Chad	861	17	-844	(-98%)
China	21597	30291	+8694	(+40%)
Colombia	100541	27018	-73523	(-73%)
Congo, Dem. Rep.	230509	199389	-31120	(-14%)
Congo, Rep.	34477	33530	-947	(-3%)
Costa Rica	5659	2653	-3006	(-53%)
Cote d'Ivoire	30012	5965	-24047	(-80%)
Cuba	14109	5786	-8323	(-59%)
Dominican Republic	5481	3381	-2100	(-38%)
Ecuador	20067	16909	-3158	(-16%)
El Salvador	2331	1369	-962	(-41%)
Equatorial Guinea	2933	2675	-258	(-9%)
Ethiopia	39590	41674	+2084	(+5%)
French Guiana	8592	4058	-4534	(-53%)
Gabon	27151	26178	-973	(-4%)
Ghana	13116	770	-12346	(-94%)
Guatemala	10667	6718	-3949	(-37%)
Guinea	18599	8595	-10004	(-54%)
Guinea-Bissau	1348	0	-1348	(-100%)
Guyana	21534	2672	-18862	(-88%)
Haiti	3382	754	-2628	(-78%)
Honduras	11616	7657	-3959	(-34%)
India	34293	18979	-15314	(-45%)
Indonesia	213246	104505	-108741	(-51%)
Jamaica	1176	0	-1176	(-100%)
Japan	164	39	-125	(-76%)
Kenya	20565	17816	-2749	(-13%)
Lao PDR	22976	12597	-10379	(-45%)
Lesotho	0	268	+268	(new)
Liberia	9908	9293	-615	(-6%)
Madagascar	53116	48359	-4757	(-9%)
Malawi	8475	4583	-3892	(-46%)
Malaysia	35720	14051	-21669	(-61%)
Mexico	54345	34878	-19467	(-36%)
Mozambique	62931	35267	-27664	(-44%)
Myanmar	47616	31854	-15762	(-33%)
Nepal	0	3312	+3312	(new)
Nicaragua	12968	7975	-4993	(-39%)
Nigeria	28792	1717	-27075	(-94%)
Panama	8959	4500	-4459	(-50%)
Papua New Guinea	51723	25520	-26203	(-51%)
Paraguay	29591	10163	-19428	(-66%)
Peru	74724	24262	-50462	(-68%)
Philippines	38768	15724	-23044	(-59%)

Country	Baseline (1000 Ha)	A2 2050 (1000 Ha)	Change (1000 Ha)	%
Rwanda	2296	2486	+190	(+8%)
Senegal	411	0	-411	(-100%)
Sierra Leone	7631	2006	-5625	(-74%)
Solomon Islands	5322	2921	-2401	(-45%)
South Africa	6796	14140	+7344	(+108%)
South Sudan	24980	3986	-20994	(-84%)
Sri Lanka	6507	1144	-5363	(-82%)
Sudan	272	0	-272	(-100%)
Suriname	14810	447	-14363	(-97%)
Swaziland	1516	708	-808	(-53%)
Tanzania UR	82305	61876	-20429	(-25%)
Thailand	34606	5546	-29060	(-84%)
Timor-Leste	1803	1242	-561	(-31%)
Togo	4082	409	-3673	(-90%)
Uganda	21578	21028	-550	(-3%)
United States of America	3960	7218	+3258	(+82%)
Venezuela	83978	12959	-71019	(-85%)
Viet Nam	27213	11551	-15662	(-58%)
Zambia	56972	39304	-17668	(-31%)
Zimbabwe	3194	992	-2202	(-69%)

0.6.7 Suitability conditional distributions

Soils and nutrients

Coffee is very sensitive to soil conditions. The Harmonized World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) contains six soil components for both the topsoil and subsoil, to study this. The comparison between the distribution across the entire tropics, and across coffee regions for Arabica farms is shown in figures 26 and 27.

From the first figure, coffee is more common in soils that have a larger share of sand and smaller share of silt than the norm. Clay also shows effects where coffee is less frequently grown in regions with intermediate quantities of clay. From the second figure, it appears that coffee is suitable in regions with intermediate quantities of calcium carbonate and low levels of gypsum.

Elevation

The distributional analysis shows a very wide range of elevations, possibly reflecting inaccuracies in the maps of Arabica and Robusta cultivation. See figure 28.

Arabica shows clear diminished presence at low elevations (below 550 m) and increased presence at all higher elevations. However, there is still probability mass below 550 m. Similarly, Robusta has extra presence of elevations below 50 m, but still has some elevated presence between 550 m and 1200 m.

The most important result of elevation for coffee cultivation is the temperatures it produces. Hawaii, for example, has excellent coffee-growing temperatures from sea level to 610m, and Arabica coffee is grown across this entire range (Thurston et al., 2013). However, the distributions shown in figure 28 are probably much more broad than is accurate. This data problem does not undermine the method, except that it increases the amount of uncertainty in the results.

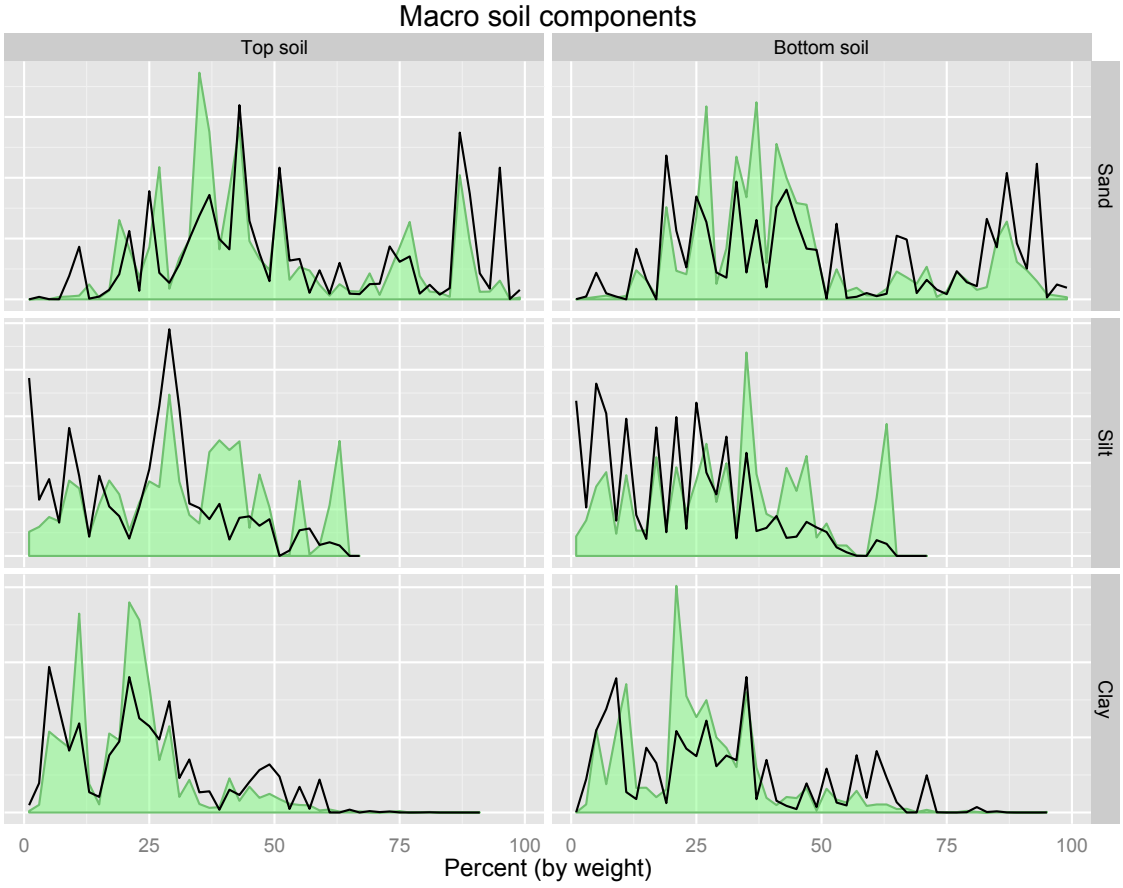


Figure 26: Comparison of distributions of texture soil components. The faded area shows the distribution of soils generally between 30°N and 30°S. The line shows the distribution of soils, weighted by coffee planting densities.

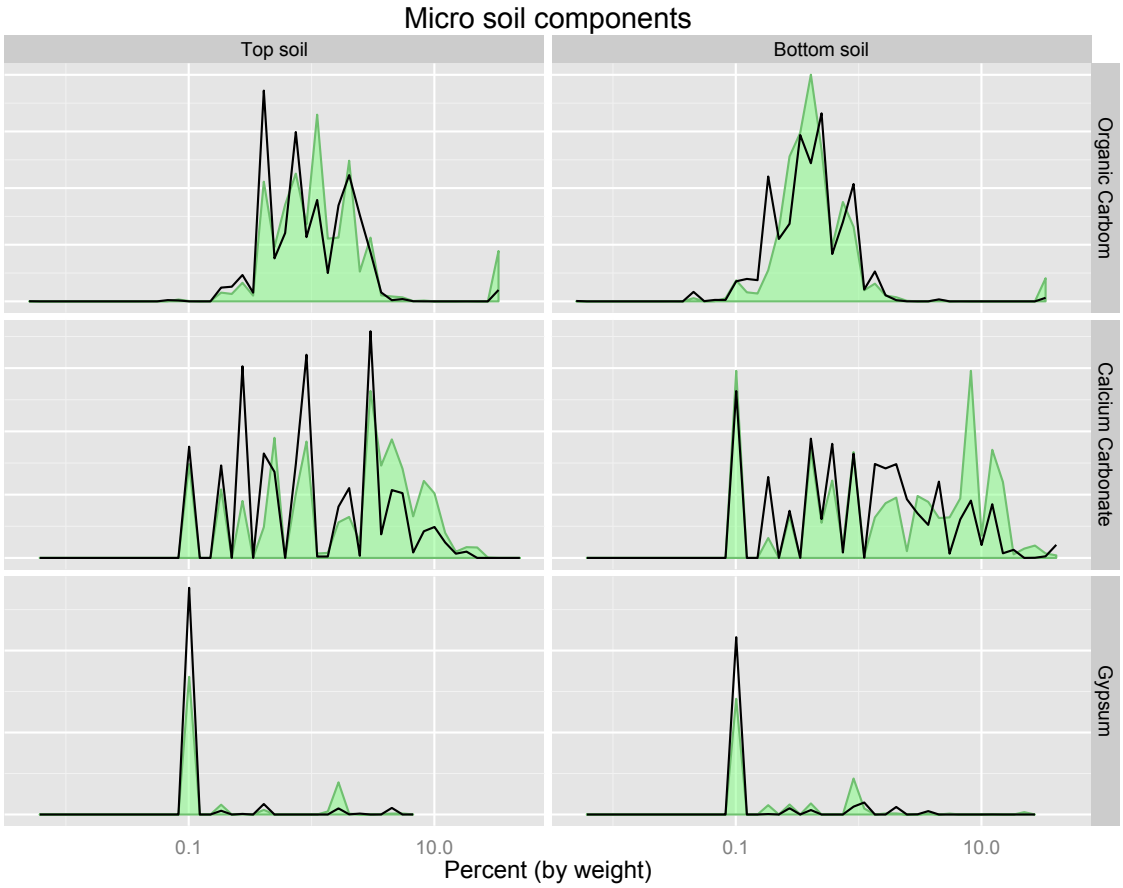


Figure 27: Comparison of distributions of trace soil components. The faded area shows the distribution of soils generally between 30°N and 30°S. The line shows the distribution of soils, weighted by coffee planting densities.

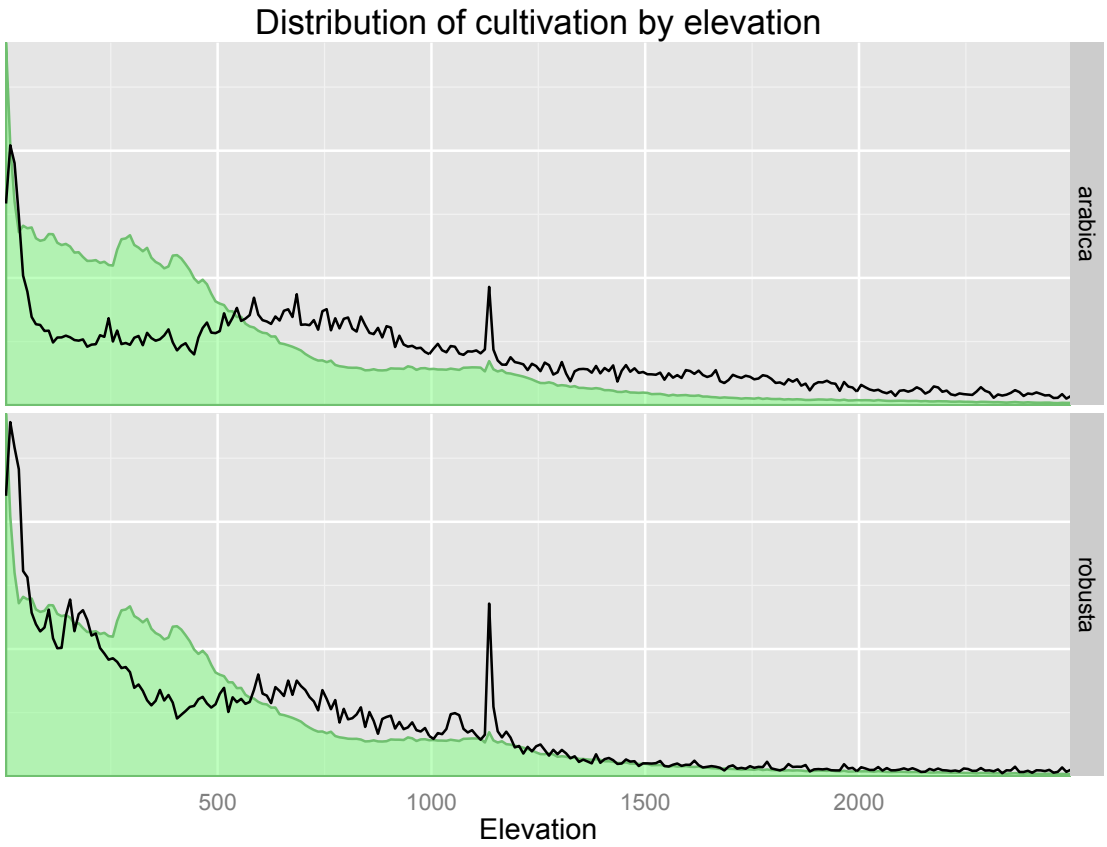


Figure 28: Distributions of elevation for Arabica and Robusta (lines) and for the tropics in general (green).

Bioclimatic variables

Figure 29 shows the distributions for all bioclimatic variables. These distributions are more erratic, because of the uneven spread of the observations within them: several bins in these histograms have no locations within their range, because of the discrete valuation of the Bioclim variables.

Latitude

We also incorporate latitude itself (see figure 31). Even if there are increases in temperature, different latitudes will provide different levels of suitability, because of the tilt of the Earth and other processes. We cannot be certain whether coffee will grow effectively outside of these latitudes, even if they appear climatically similar in the future to lower latitudes now. Including the distribution of latitude imposes a slight conservativism on our estimate which is supported by the data.

0.6.8 Changes in suitability by country for our model

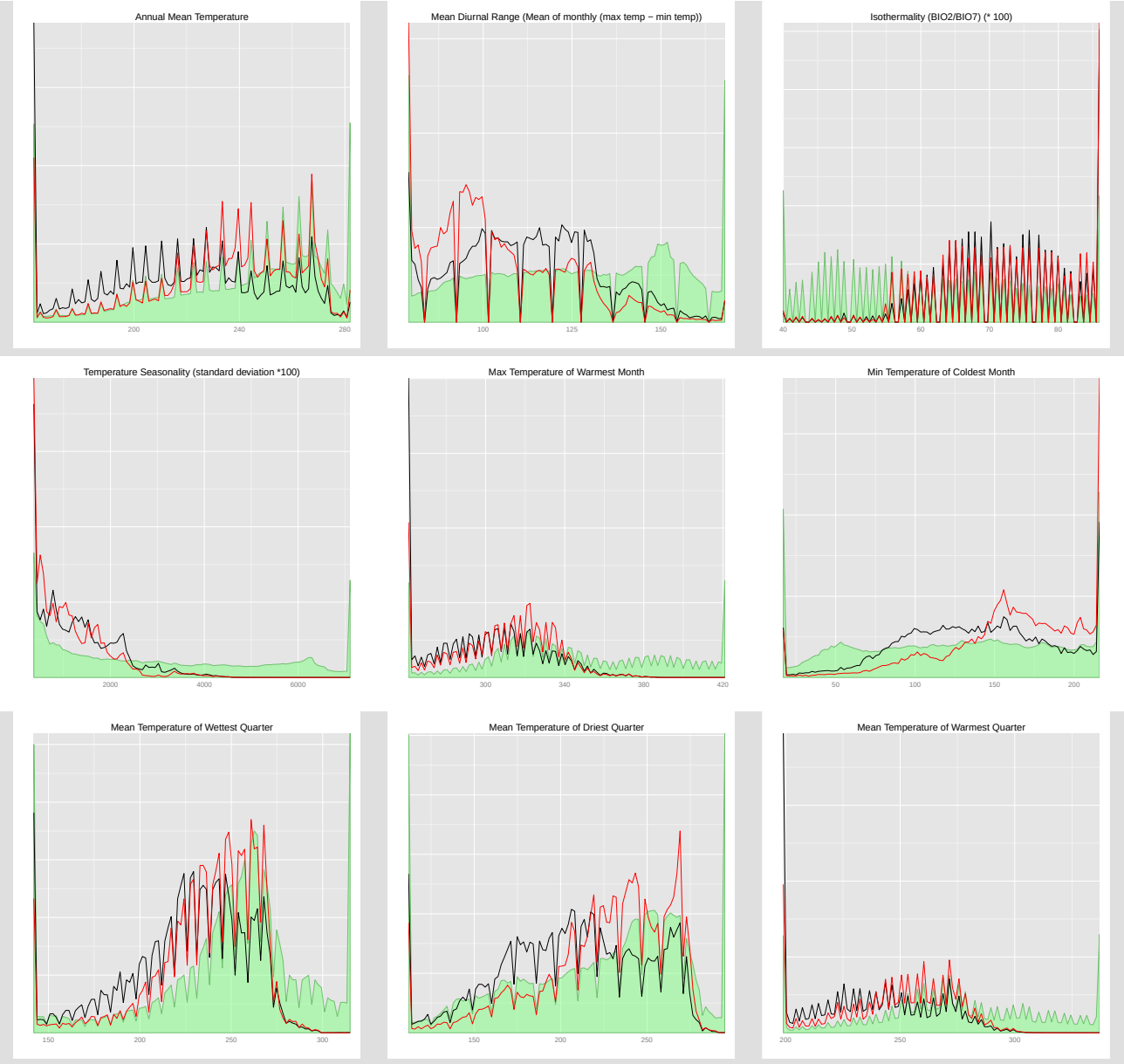


Figure 29: First set of nine of the 19 variables in the Bioclim dataset, with coffee region distributions shown in black (Arabica) and red (Robusta). We dropped one, the Annual Temperature Range, since the technique implicitly infers it from the minimum and maximum temperatures.

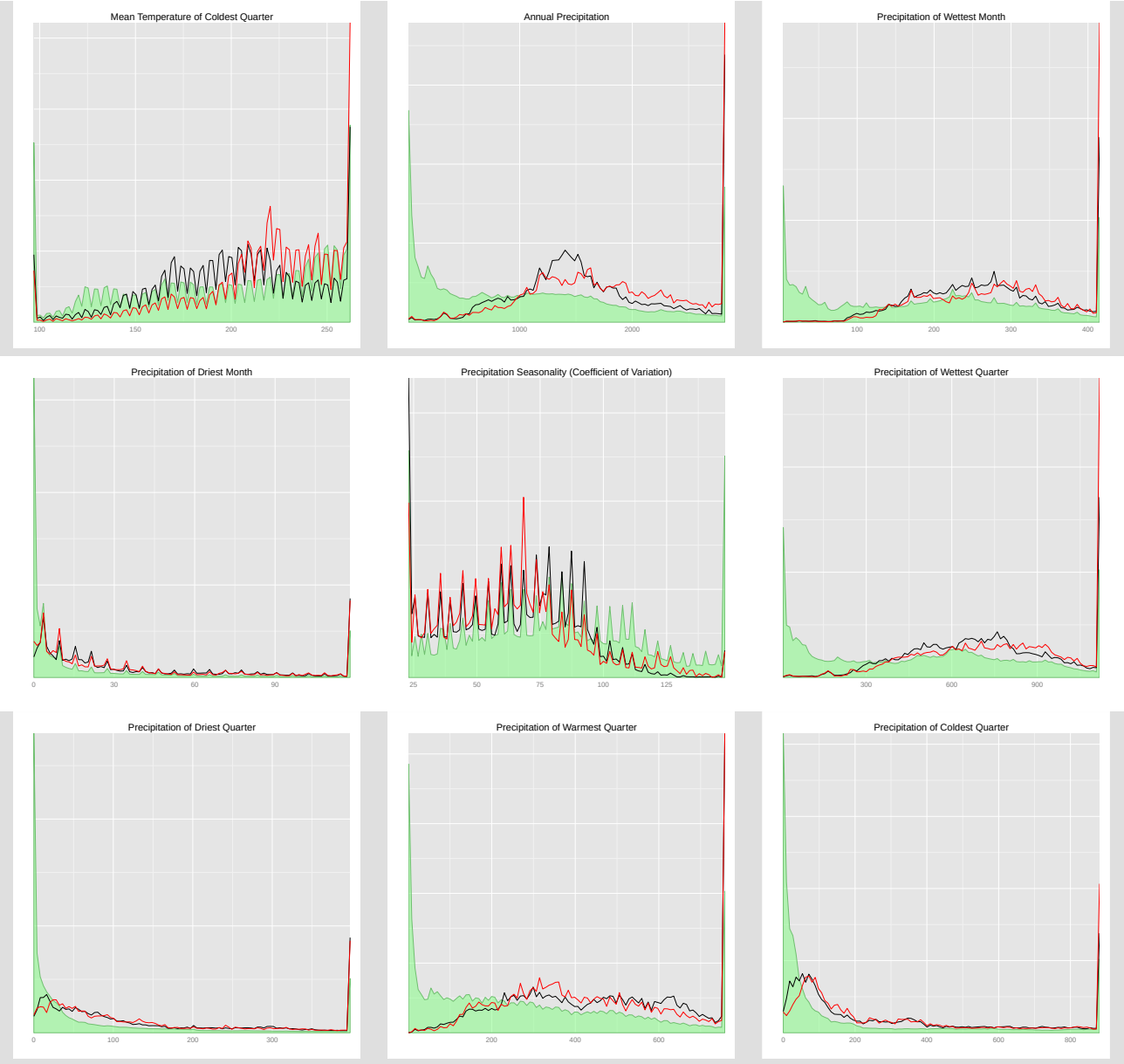


Figure 30: Second set of nine of the 19 variables in the Bioclim dataset, with coffee region distributions shown in black (Arabica) and red (Robusta). We dropped one, the Annual Temperature Range, since the technique implicitly infers it from the minimum and maximum temperatures.

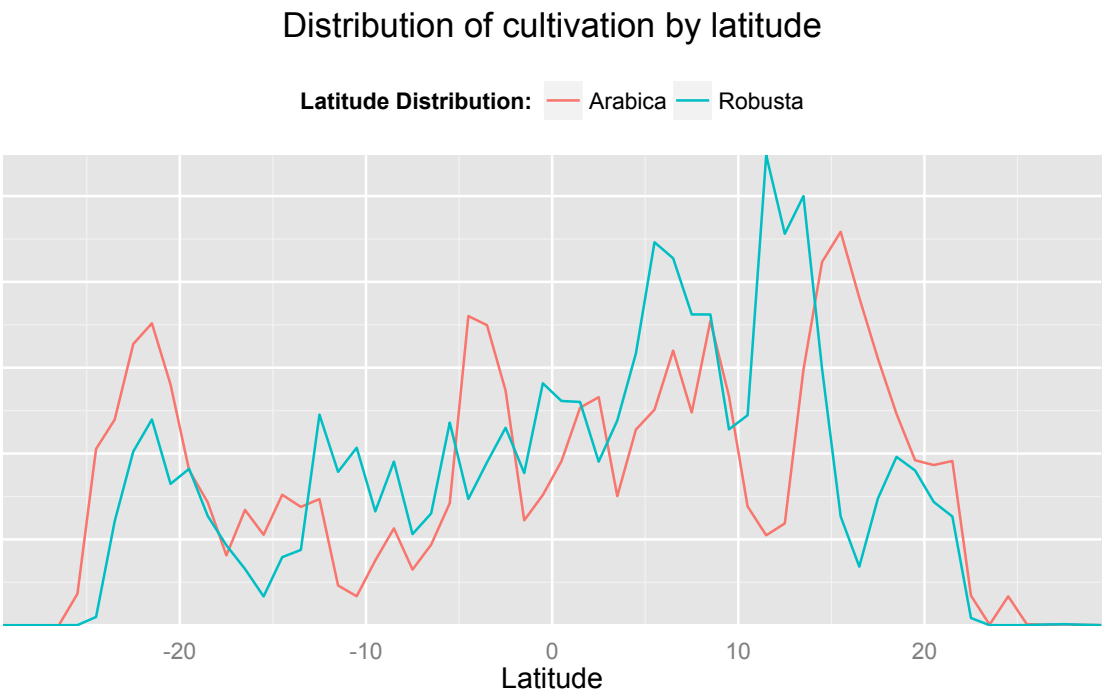


Figure 31: The distribution of coffee production for Arabica (red) and Robusta (blue) across latitude.

Arabica Changes, Baseline - 2050, RCP 8.5

	Suitable Areas						Harvested Areas			
Country	Baseline (Ha)	Increase (Ha)	Conf (%)	Decrease (Ha)	Conf. (%)	Loss (%)	Chng. (%)	Harvest (Ha)	H. Loss (%)	Conf. (%)
Angola	7935811	3205274	90.55	-7639694	99.23	-96.30	-55.90	31000	-43.90	40.84
Argentina	712904	507034	88.73	-367099	94.30	-51.50	19.60	0	0.00	
Australia	328392	0		-324389	97.77	-98.80	-98.80	0	0.00	
Burundi	651014	46827	69.36	-469684	89.31	-72.10	-65.00	27000	-25.70	28.59
Belize	29037	0		-29037	100.00	-100.00	-100.00	47	-100.00	0.44
Bolivia	3014467	366559	98.52	-861014	100.00	-28.60	-16.40	30000	-5.40	6.54
Brazil	30428040	4118211	84.74	-26059433	99.72	-85.60	-72.10	2120080	-34.70	34.80
Bhutan	0	94778	70.01	0				0	0.00	
Botswana	2594	0		-2594	100.00	-100.00	-100.00	0	-100.00	
Central African Republic	17244	0		-17244	100.00	-100.00	-100.00	14000	-100.00	0.05
Chile	332	245274	99.17	-331	52.94	-99.80		0	0.00	
China	271465	32363	62.09	-254134	94.85	-93.60	-81.70	62000	-37.90	36.73
Cote d'Ivoire	34712	0		-34712	100.00	-100.00	-100.00	160000	-100.00	0.79
Cameroon	913518	16185	87.87	-645552	100.00	-70.70	-68.90	210000	-25.50	26.61
Democratic Republic of the Congo	9173840	332384	56.07	-7480138	97.31	-81.50	-77.90	86000	-29.40	36.54
Republic of Congo	277884	0		-277884	100.00	-100.00	-100.00	10500	-100.00	34.36
Colombia	6417055	1145273	93.02	-3377767	95.02	-52.60	-34.80	778084	-19.30	23.09
Comoros	52502	612	50.00	-40458	96.81	-77.10	-75.90	885	0.00	
Cape Verde	560	8291	100.00	-560	91.13	-100.00		0	-100.00	
Costa Rica	689981	126846	99.53	-393518	96.60	-57.00	-38.60	93774	-14.70	20.06
Cuba	157527	0		-157441	98.11	-99.90	-99.90	28000	-7.40	7.29
Dominican Republic	214847	48433	77.45	-65053	100.00	-30.30	-7.70	133342	-2.10	2.86
Ecuador	2197675	294876	86.17	-831263	92.68	-37.80	-24.40	78709	-17.50	19.95
Eritrea	624377	11328	93.56	-449558	100.00	-72.00	-70.20	0	0.00	
Spain	0	146972	97.54	0				0	0.00	
Ethiopia	10371465	2145052	80.89	-6924854	96.70	-66.80	-46.10	528571	-28.00	26.38
Fiji	28692	0	17.61	-28692	100.00	-100.00	-100.00	30	-100.00	
France	140087	33214	75.23	-45928	100.00	-32.80	-9.10	120	0.00	
Gabon	9576	0		-9576	100.00	-100.00	-100.00	310	-100.00	0.64
Guinea	102460	0		-102460	100.00	-100.00	-100.00	66000	-100.00	8.79
Equatorial Guinea	143859	0	5.88	-143859	100.00	-100.00	-100.00	12500	-100.00	25.83
Guatemala	2045755	291340	76.44	-947309	94.23	-46.30	-32.10	250000	-25.70	28.45
Guyana	703847	1525	41.53	-673630	99.81	-95.70	-95.50	360	-5.20	5.23
Honduras	2582984	63693	67.23	-1614594	100.00	-62.50	-60.00	266000	-22.90	29.53
Haiti	340896	18112	73.09	-211364	100.00	-62.00	-56.70	92000	-27.00	28.04
Indonesia	7718460	715065	85.38	-4967199	98.11	-64.40	-55.10	1233900	-10.90	12.85
India	530599	135216	92.02	-333350	98.50	-62.80	-37.30	368687	-1.00	1.06
Jamaica	49656	0		-43093	100.00	-86.80	-86.80	7500	-19.80	23.34
Kenya	5174101	899779	79.31	-3242596	95.39	-62.70	-45.30	160000	-27.10	29.89
Cambodia	2270	0		-2270	100.00	-100.00	-100.00	430	-100.00	0.02
Lao PDR	219546	0	83.04	-214642	100.00	-97.80	-97.80	56875	-19.00	17.61
Sri Lanka	49459	1921	77.70	-26821	98.61	-54.20	-50.30	8460	-0.70	0.36
Lesotho	0	1645	76.47	0				0	0.00	
Madagascar	15883957	1377660	89.64	-7587513	97.72	-47.80	-39.10	138000	-18.90	22.64

Arabica Changes, Baseline - 2050, RCP 8.5

	Suitable Areas						Harvested Areas			
Country	Baseline (Ha)	Increase (Ha)	Conf (%)	Decrease (Ha)	Conf. (%)	Loss (%)	Chng. (%)	Harvest (Ha)	H. Loss (%)	Conf. (%)
Mexico	23350727	3808753	88.87	-12412932	98.49	-53.20	-36.80	695350	-21.10	24.37
Myanmar	83638	4909	41.16	-80280	96.62	-96.00	-90.10	12000	-17.60	17.85
Mozambique	3412646	145274	52.03	-3013822	98.65	-88.30	-84.10	980	-18.40	16.78
Mauritius	64299	0		-48778	100.00	-75.90	-75.90	0	0.00	
Malawi	1566109	260273	57.43	-1289161	96.59	-82.30	-65.70	2580	-1.10	1.03
Malaysia	569169	20010	77.03	-498527	96.93	-87.60	-84.10	19300	-0.50	0.47
Namibia	1225713	135934	75.18	-1187122	99.08	-96.90	-85.80	0	0.00	
New Caledonia	117779	4	22.00	-101959	99.37	-86.60	-86.60	95	0.00	
Nigeria	3436	111	41.52	-3417	93.87	-99.40	-96.20	2200	-2.20	1.97
Nicaragua	666380	10090	67.91	-552274	100.00	-82.90	-81.40	123000	-13.70	16.48
Nepal	0	0	63.34	0				1780	0.00	0.00
Oman	0	0		0.00	100	-100.00	-100.00	0.00	-100	
Panama	384634	40553	86.14	-224428	96.32	-58.30	-47.80	30000	-20.30	27.71
Peru	3297382	2784251	90.64	-1746405	95.76	-53.00	31.50	312251	-13.80	12.72
Philippines	1471048	104253	85.25	-1009926	98.86	-68.70	-61.60	119999	-5.40	5.38
Papua New Guinea	4713564	278522	85.31	-2162547	97.79	-45.90	-40.00	73000	-10.60	20.15
Puerto Rico	150857	0		-124692	100.00	-82.70	-82.70	42000	-36.70	46.40
Paraguay	0	0		0	100.00	-100.00	-100.00	300	-100.00	0.00
Rwanda	870139	110162	91.99	-580954	88.53	-66.80	-54.10	41762	-9.20	15.87
Saudi Arabia	813254	1126	92.77	-778853	99.04	-95.80	-95.60	0	0.00	
South Sudan	142106	5689	78.29	-112004	94.08	-78.80	-74.80	0	0.00	
Solomon Islands	45500	6006	88.23	-45464	100.00	-99.90	-86.70	0	0.00	
Sierra Leone	0	0		0.00	100	-100.00	-100.00	14000.00	-100	0.05
El Salvador	89400	11899	79.65	-60542	92.26	-67.70	-54.40	139958	-33.90	34.11
Somaliland	542582	11131	47.56	-540533	100.00	-99.60	-97.60	0	0.00	
Somalia	95715	4192	0.00	-95715	99.99	-100.00	-95.60	0	-100.00	
Sao Tome and Principe	23219	0		-16693	100.00	-71.90	-71.90	250	0.00	
Swaziland	95813	8141	47.56	-80438	81.81	-84.00	-75.50	0	0.00	
Thailand	239895	0	48.53	-239890	100.00	-100.00	-100.00	52000	-100.00	0.26
Timor-Leste	204294	15466	100.00	-163624	99.84	-80.10	-72.50	55000	-12.10	14.41
Taiwan	18096	45791	79.81	-18096	100.00	-100.00	153.00	0	-100.00	
Tanzania	10254620	1768496	82.71	-8586347	99.57	-83.70	-66.50	127335	-21.80	21.42
Uganda	5074034	68432	75.12	-4830939	100.00	-95.20	-93.90	310000	-60.20	58.85
United States	463542	21137	71.67	-405753	100.00	-87.50	-83.00	2550	-85.70	85.70
Venezuela	5301618	262843	73.94	-3778788	98.41	-71.30	-66.30	182000	-3.70	5.46
Vietnam	1488516	74616	95.56	-1272729	98.61	-85.50	-80.50	574314	-1.90	1.86
Vanuatu	14725	0	49.87	-14608	90.77	-99.20	-99.20	65	0.00	
Yemen	2988863	16488	89.64	-2166101	98.70	-72.50	-71.90	34987	-27.40	30.29
South Africa	2670584	770739	88.69	-2282693	100.00	-85.50	-56.60	0	0.00	
Zambia	4464733	13902	1.57	-4249238	95.99	-95.20	-94.90	7000	-34.40	33.27
Zimbabwe	404205	6145	49.40	-395711	100.00	-97.90	-96.40	5397	-31.40	25.28

Robusta Changes, Baseline - 2050, RCP 8.5

		Suitable Areas					Harvested Areas			
Country	Baseline (Ha)	Increase (Ha)	Conf (%)	Decrease (Ha)	Conf. (%)	Loss (%)	Chng. (%)	Harvest (Ha)	H. Loss (%)	Conf. (%)
Angola	6057845	6038395	84.03	-2842494	89.91	-46.90	52.80	31000	-14.90	10.66
Argentina	5619105	2133134	83.48	-1129718	100.00	-20.10	17.90	0	0.00	
Antigua and Barbuda	0	0	100.00	0		0.00		0	0.00	
Australia	8333570	3459763	93.97	-116037	69.03	-1.40	40.10	0	0.00	
Burundi	367007	147167	75.90	-288042	100.00	-78.50	-38.40	27000	-12.10	17.96
Benin	1308471	0		0		0.00	0.00	350	0.00	0.00
Bangladesh	550782	568907	15.07	0		0.00	103.30	0	0.00	
Bahamas	602877	0	50.00	0		0.00	0.00	0	0.00	
Belize	1135507	897102	82.07	0		0.00	79.00	47	0.00	0.00
Bolivia	12425186	9041991	61.06	-975638	100.00	-7.90	64.90	30000	-0.00	0.14
Brazil	160377693	36056814	71.65	-2334785	77.78	-1.50	21.00	2120080	-11.70	8.44
Barbados	49673	12	100.00	0		0.00	0.00	0	0.00	
Brunei Darussalam	297062	13	100.00	0		0.00	0.00	0	0.00	
Bhutan	0	105732	99.09	0				0.00	0	
Botswana	1857	2417951	55.89	-410520	100.00	-22104.30		0	0.00	
Central African Republic	29832739	3647129	85.28	-33843	37.67	-0.10	12.10	14000	-0.00	0.00
Chile	0	6	0.00	-13108	100.00	-Inf	-Inf	0.00	0.00	
China	2311341	557426.00	83.21	-3512451.00	100.00	-152.00	-127.80	62000.00	-2.30	6.26
Cote d'Ivoire	9342887	1236486	94.35	-301	29.41	-0.00	13.20	160000	-0.00	0.02
Cameroon	17549050	6070796	97.80	-178030	64.39	-1.00	33.60	210000	-1.10	2.28
Democratic Republic of the Congo	37231640	16218451	82.71	-6834249	82.75	-18.40	25.20	86000	-0.60	0.58
Republic of Congo	4995468	2556223	96.36	-359036	67.98	-7.20	44.00	10500	-5.40	3.95
Colombia	22966693	6251581	70.99	-2758247	95.20	-12.00	15.20	778084	-1.50	8.71
Comoros	30354	41543	78.33	0		0.00	136.90	885	0.00	
Cape Verde	49443	10263	92.50	-18158	79.60	-36.70	-16.00	0	0.00	
Costa Rica	943401	160983	77.69	-274849	82.03	-29.10	-12.10	93774	-7.60	6.74
Cuba	6011703	1518611	83.70	-381	100.00	-0.00	25.30	28000	-0.00	1.55
Cayman Islands	0	15292	100.00	0		0.00		0	0.00	
Djibouti	8378	0	88.23	0		0.00	0.00	0	0.00	
Dominica	75759	18386	100.00	0		0.00	24.30	610	0.00	
Dominican Republic	1038123	1211488	92.47	-74612	89.21	-7.20	109.50	133342	-0.30	1.55
Ecuador	5692491	1070520	85.31	-1284708	95.13	-22.60	-3.80	78709	-1.10	1.75
Egypt	7184	0		0		0.00	0.00	0	0.00	
Eritrea	1153440	93699	100.00	-307887	94.96	-26.70	-18.60	0	0.00	
Spain	0	55278	79.99	0				0.00	0	
Ethiopia	7311617	2624156	83.06	-6071007	93.75	-83.00	-47.10	528571	-13.90	25.73
Fiji	727009	218796	91.87	-1709	44.03	-0.20	29.90	30	0.00	
France	3229097	97349	99.84	-15076	98.98	-0.50	2.50	120	0.00	
Gabon	9875969	5485485	99.64	-9440	34.26	-0.10	55.40	310	-0.00	0.00
Ghana	8032693	198856	98.17	0		0.00	2.50	4000	0.00	0.00
Guinea	8170406	842479	90.84	-246659	77.00	-3.00	7.30	66000	-6.00	4.53
Guinea-Bissau	160378	0		0		0.00	0.00	0	0.00	
Equatorial Guinea	191644	322974	99.57	-12	33.19	-0.00	168.50	12500	-0.10	0.04
Grenada	33577	23514	97.94	0		0.00	70.00	0	0.00	

Robusta Changes, Baseline - 2050, RCP 8.5

	Suitable Areas						Harvested Areas			
Country	Baseline (Ha)	Increase (Ha)	Conf (%)	Decrease (Ha)	Conf. (%)	Loss (%)	Chng. (%)	Harvest (Ha)	H. Loss (%)	Conf. (%)
Guatemala	3618767	2687814	79.36	-887998	92.36	-24.50	49.70	250000	-8.20	31.09
Guyana	4484150	425361	84.95	-48908	23.99	-1.10	8.40	360	-0.30	0.43
Hong Kong	0	14805	74.94	0				0.00	0	
Honduras	6825530	2312673	93.64	-711645	83.06	-10.40	23.50	266000	-9.80	12.79
Haiti	1664408	789640	92.24	-29286	87.31	-1.80	45.70	92000	-0.10	1.12
Indonesia	74982682	19663507	94.72	-1724781	73.65	-2.30	23.90	1233900	-2.10	3.38
India	8280081	2222490	92.47	-85615	88.18	-1.00	25.80	368687	-1.70	2.30
Jamaica	162722	61860	92.30	0		0.00	38.00	7500	0.00	0.00
Kenya	8225913	1255119	82.00	-2540490	95.54	-30.90	-15.60	160000	-5.90	5.69
Cambodia	7086551	432861	58.70	-24583	31.28	-0.30	5.80	430	-0.00	0.00
Saint Kitts and Nevis	3030	0	100.00	0		0.00	0.00	0	0.00	
Lao PDR	4223896	1285612	65.94	-277249	69.64	-6.60	23.90	56875	-3.00	0.78
Liberia	3450601	1341275	99.43	0		0.00	38.90	2800	0.00	0.00
Saint Lucia	10554	106	100.00	0		0.00	1.00	0	0.00	
Sri Lanka	1013890	44733	95.99	-37293	77.70	-3.70	0.70	8460	-0.00	0.05
Morocco	3502681	1979440	87.06	0		0.00	56.50	0	0.00	
Madagascar	5852964	5983642	93.51	-435926	99.51	-7.40	94.80	138000	-12.70	12.69
Mexico	18029553	9691106	81.94	-16710010	100.00	-92.70	-38.90	695350	-7.50	25.78
Myanmar	2268399	985504	87.47	-702709	100.00	-31.00	12.50	12000	-0.00	0.02
Mozambique	7928756	8212801	86.37	-472518	94.03	-6.00	97.60	980	-8.20	8.63
Mauritania	322224	9265	99.83	0		0.00	2.90	0	0.00	
Montserrat	8224	0	88.23	0		0.00	0.00	0	0.00	
Mauritius	86175	108848	89.97	0	100.00	-0.00	126.30	0	0.00	
Malawi	3217741	1174015	82.66	-262941	81.31	-8.20	28.30	2580	-9.50	11.93
Malaysia	11411462	2194454	95.78	-28954	65.05	-0.30	19.00	19300	-0.10	0.05
Namibia	70533	194978	81.62	-1317228	100.00	-1867.50	-1591.10	0	0.00	
New Caledonia	336290	604784	99.88	-1103	96.29	-0.30	179.50	95	0.00	
Nigeria	10354095	128681	89.50	-2205	68.58	-0.00	1.20	2200	-0.20	0.11
Nicaragua	6028246	1797003	97.00	-54329	63.18	-0.90	28.90	123000	-0.00	0.06
Oman	168703	83281	90.57	0		0.00	49.40	0	0.00	
Panama	1724571	545477	88.33	-165258	68.57	-9.60	22.00	30000	-5.10	8.91
Peru	11478256	6664944	72.28	-4432821	93.63	-38.60	19.40	312251	-3.80	9.95
Philippines	9003144	3005544	89.84	-372913	81.48	-4.10	29.20	119999	-4.30	4.86
Papua New Guinea	20078001	4528851	93.59	-1292919	82.69	-6.40	16.10	73000	-4.10	11.47
Puerto Rico	537072	356958	98.38	0		0.00	66.50	42000	0.00	0.00
Paraguay	3755829	2516351	84.98	0	64.71	-0.00	67.00	300	-0.00	0.01
Rwanda	206056	8233	43.14	-519559	93.52	-252.10	-248.10	41762	-4.60	37.08
Western Sahara	40247	32087	99.03	0		0.00	79.70	0	0.00	
Saudi Arabia	1065006	46629	99.99	-304052	100.00	-28.50	-24.20	0	0.00	
Sudan	371759	19678	97.53	0		0.00	5.30	0	0.00	
South Sudan	3433923	310336	58.98	-31977	71.12	-0.90	8.10	0	0.00	
Solomon Islands	772505	383510	98.42	-63	100.00	-0.00	49.60	0	0.00	
Sierra Leone	2200081	490018	88.94	-2319	66.23	-0.10	22.20	14000	-0.00	0.04
El Salvador	406748	113872	64.22	-60727	58.55	-14.90	13.10	139958	-3.70	3.00

Robusta Changes, Baseline - 2050, RCP 8.5

	Suitable Areas						Harvested Areas			
Country	Baseline (Ha)	Increase (Ha)	Conf (%)	Decrease (Ha)	Conf. (%)	Loss (%)	Chng. (%)	Harvest (Ha)	H. Loss (%)	Conf. (%)
Somaliland	1453739	407729	98.39	-163611	46.73	-11.30	16.80	0	0.00	
Somalia	2096908	58334	45.79	0	68.45	-0.00	2.80	0	0.00	
Sao Tome and Principe	54718	11894	100.00	-177	23.53	-0.30	21.40	250	0.00	
Suriname	3071143	6817	31.93	0		0.00	0.20	488	0.00	0.00
Swaziland	75788	15519	100.00	-16263	100.00	-21.50	-1.00	0	0.00	
Togo	1776596	101957	100.00	0		0.00	5.70	27000	0.00	0.00
Thailand	9957365	648750	82.92	-71400	54.30	-0.70	5.80	52000	-0.00	0.00
Timor-Leste	163147	147219	96.52	-8871	35.65	-5.40	84.80	55000	-0.90	1.73
Tonga	35643	7387	96.30	0		0.00	20.70	10	0.00	
Trinidad and Tobago	160920	33	50.01	0		0.00	0.00	350	0.00	0.00
Taiwan	373110	260304	90.01	0	41.19	-0.00	69.80	0	0.00	
Tanzania	17851110	5624377	89.30	-8989117	83.21	-50.40	-18.80	127335	-12.40	15.03
Uganda	10507129	1174475	88.39	-5832694	79.42	-55.50	-44.30	310000	-58.40	46.56
United States	601889	364248	88.34	-30276	83.89	-5.00	55.50	2550	-22.00	24.19
Saint Vincent and the Grenadines	133	74	100.00	0		0.00	55.40	0	0.00	
Venezuela	1391729	2344465	74.74	-2256727	81.41	-16.90	0.70	182000	-2.20	2.49
United States Virgin Islands	586	0	100.00	0		0.00	0.00	0	0.00	
Vietnam	11189490	4400852	88.86	-93588	88.27	-0.80	38.50	574314	-0.30	3.90
Vanuatu	677195	331861	97.18	-1027	27.53	-0.20	48.90	65	0.00	
Samoa	44108	11554	58.88	0		0.00	26.20	35	0.00	
Yemen	6053411	2158155	88.47	-998765	99.37	-16.50	19.20	34987	-2.40	58.69
South Africa	1169832	722635	97.66	-4688250	100.00	-400.80	-339.00	0	0.00	
Zambia	3234348	713801	78.20	-1077771	91.74	-33.30	-11.30	7000	-52.70	41.85
Zimbabwe	719338	352388	56.14	-968593	100.00	-134.70	-85.70	5397	-12.60	15.06

0.7 Extra variability analysis

0.7.1 Computing ENSO impacts

We estimate the impacts of El Niño and La Niña by estimating an "impulse response", which accounts for the multiple overlapping effects of different ENSO years and the monthly climatology of the NINO 3.4 signal.

$$y_t = \alpha + \sum_{Y=Year(t)-N/12+1}^{Year(t)} \sum_{M=1}^{N} \beta_{12(Year(t)-Y)+M}^{Class(Y)} + \gamma \sum_{s=1}^{24} \frac{y_{t-s}}{24} + \mu_{Month(t)}$$

$Year(t)$ is the year for time t and $Month(t)$ is the month for time t; $Class(Y)$ is the class of ENSO event that happened in year Y (El Niño and La Niña). N is the number of months to include in the impulse responses.

Here, the β_m^k variables describe impulse responses of length N for each class of ENSO event.

As a diagnostic of both the technique and the years used, we apply this impulse response technique to the NINO 3.4 index itself. The NINO 3.4 index is a measure of the strength of the current state of the ENSO cycle, and we can use these year definitions to construct the characteristic evolution of the NINO 3.4 during each of these years. See figure 32.

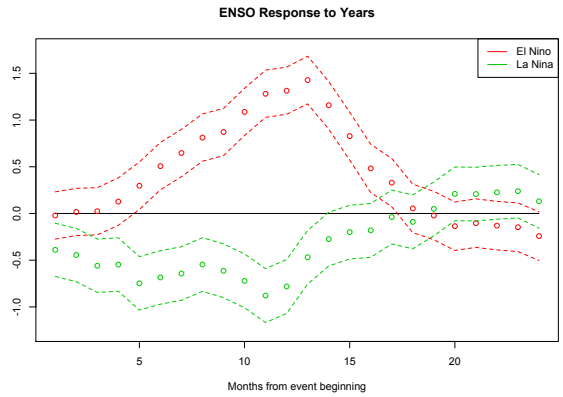


Figure 32: The estimated 24-month impulse response of the NINO 3.4 indicator to each of the three ENSO year types.

0.7.2 Additional PCA details

0.7.3 Monthly production

Production records are generally maintained on a yearly basis, but different price information is available monthly. Different countries harvest and ship beans during different months, and this information can be used to infer the monthly production added to the global market.

We use the coffee harvested calendar from the Sweet Maria's Coffee Production Timetable, which is admittedly uncertain and subject to yearly change. However, they provide a general cycle around which

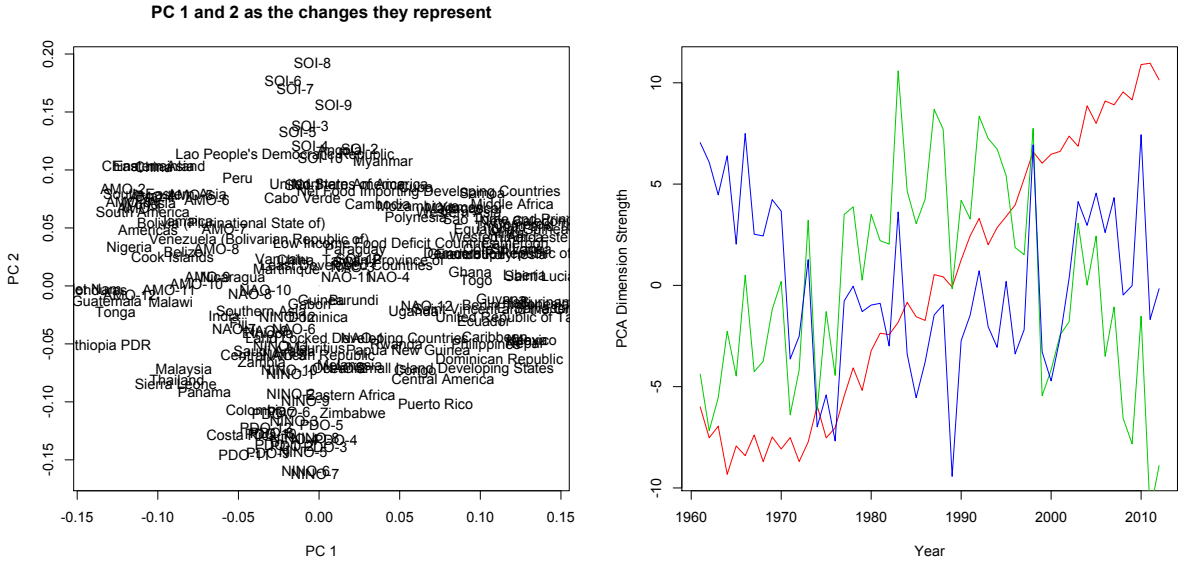


Figure 33: **Left:** The first and second principal component, in terms of the marginal effects of countries and climate signals. These are displayed more clearly in the main report. **Right:** The values of the first three principal components (PC 1 = red, PC 2 = green, PC 3 = blue) across years. As years progress, PC 1 generally increases, and PC 2 first decreases and then increases.

actual yearly production is assumed to vary. We distribute the production for each country amongst its harvesting months, and evenly distribute throughout the year production for countries not represented in the calendar (most notably, Vietnam). We also distinguish between countries that produce Arabica and Robusta coffees, or those that produce a combination of both every year. The result in the figure above shows wide variations from month to month.

It is also interesting that the range of variation has increased significantly. The peak of production each year has increased much faster than the yearly minimum: In the 1960s, the best years produced monthly peaks over 10 million bags, while the slowest months produced only 3 million bags. In the last decade, the greatest monthly production has been over 15 million bags, but the worst months have only produced 5 million bags. The situation is even starker for Arabica coffee, where the worst months in the last decade are comparable to those in the 1960s, although the best months have increased over 20%.

We can use this monthly production data to inform the coffee market model, described below.

0.7.4 Coherent movements

The relatively weak statistical relationship found between El Niño and country-specific yields is not uncommon among agricultural crops, but it drove an interest in our group into dissecting more clearly the relationship between global climate signals and country production. We collected five oceanic signals to explore this further, as shown in figure 1.6.[3]

[3]NINO 3.4, NAO, SOI, PDO from NOAA Climate Prediction Center (CPC) (2015), unsmoothed AMO from Enfield et al. (2001).

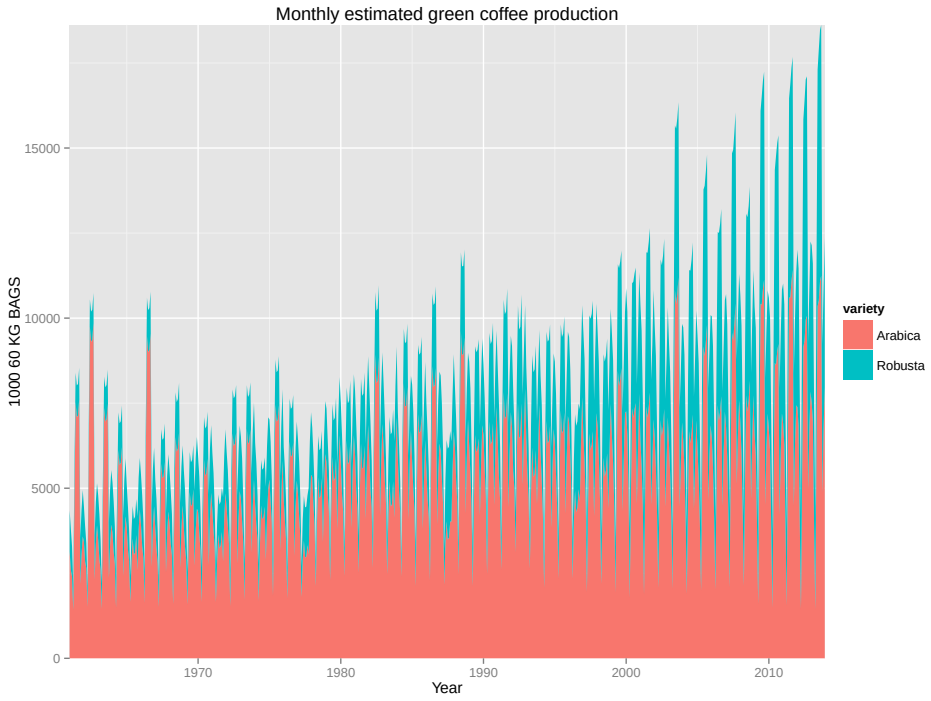


Figure 34: Inferred monthly production for Arabica and Robusta coffee, based on the harvesting calendars of their producing countries.

A principal component analysis identifies regions of coherent marginal changes, across multiple time-series. This technique can be used to better understand patterns in large datasets, like the one describing country coffee production.

For each year, monthly values of the climate signals (delayed 6 months, to capture their impacts on coffee flowering) and country yields (detrended with locfit and normalized) are included. The first principal component represents the largest coherent movement of change, followed by the second component, and so on. Between the first three components, over 50% of the variation in yields can be described. The share of each of these components by year is shown in Appendix 0.7.2. Each of the components and what it suggests about the relationship between climate and yields is described below.

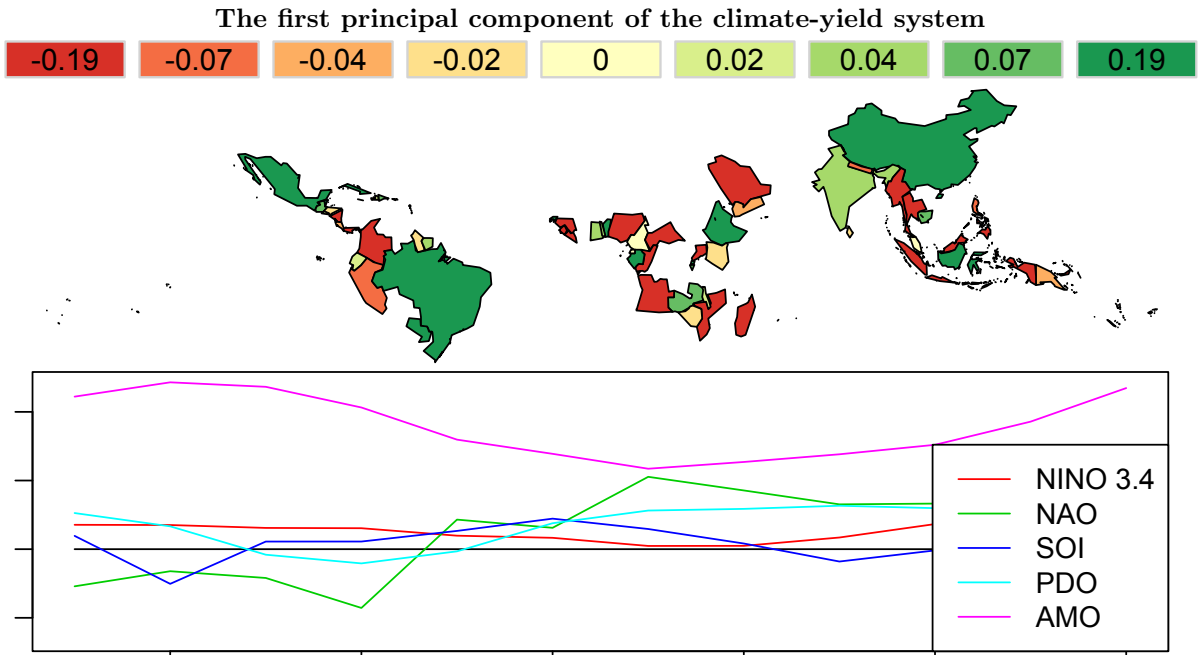


Figure 35: Spatial and temporal representation of the first principal component of the climate-yield system. Colors in the map represent increases (green) and decreases (red), and the plot below shows the climate signals across the year (delayed 6 months, so month 1 is July and month 12 is June). Explanation in the text.

The first principal component describes how yields have shifted on average over the past 50 years. Brazil, Mexico, and China have seen some the largest increases in yield, while Thailand, Myanmar and many countries in Africa have experienced the largest decreases. Most climate signals have not shown any trend, except for the Atlantic multidecadal oscillation (AMO) which is currently much higher than it was in the 1960s. As a result, all of the climate signals in the lower graph are near zero, except for AMO.

The second and third principal components are dominated by ENSO (the El Niño/La Niña cycle), represented by the NINO 3.4 index and the Southern oscillation index, which is known to be strongly correlated with ENSO but with an opposite sign. PC 2 is represented in the data when NINO 3.4 is high (El Niño) and the Pacific decadal oscillation (PDO) is also high, and its effects are reversed when these signals are both opposite in the direction of their anomalies. The largest effect of this combination, as shown in the map, is that Brazil, Paraguay, and Papua New Guinea have decreases in yields while India sees increases. This suggests that yields in these regions will often move in opposite directions, during many El Niño and La Niña years.

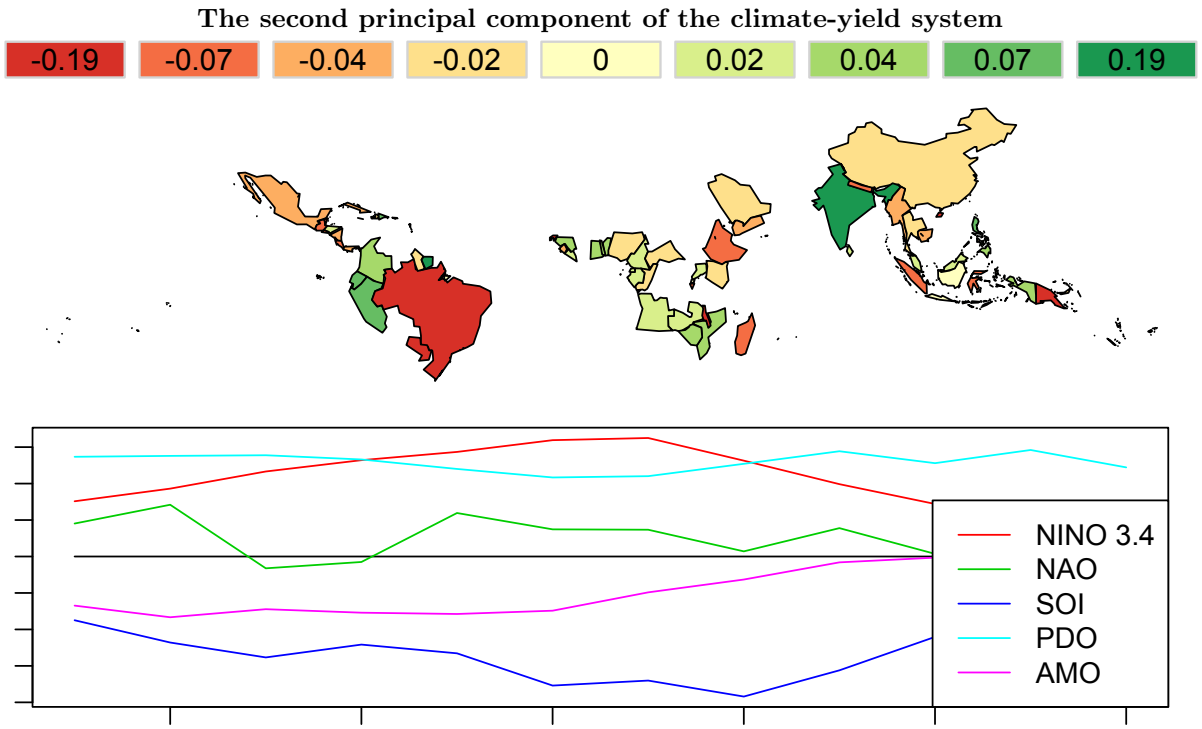


Figure 36: Spatial and temporal representation of the second principal component of the climate-yield system. Colors in the map represent increases (green) and decreases (red), and the plot below shows the climate signals across the year (delayed 6 months). Explanation in the text.

Observations with low values of PC 2 occur before 1975 and after 2000, while those with high values of PC 2 occur mostly in the 1980s and early 1990s. This may be driven by the slow oscillation of PDO. Since only one such cycle has occurred, it is difficult to distinguish the effects of the climate signals from socioeconomic effects, although most of this was be removed by the flexible trend used in the preprocessing step.

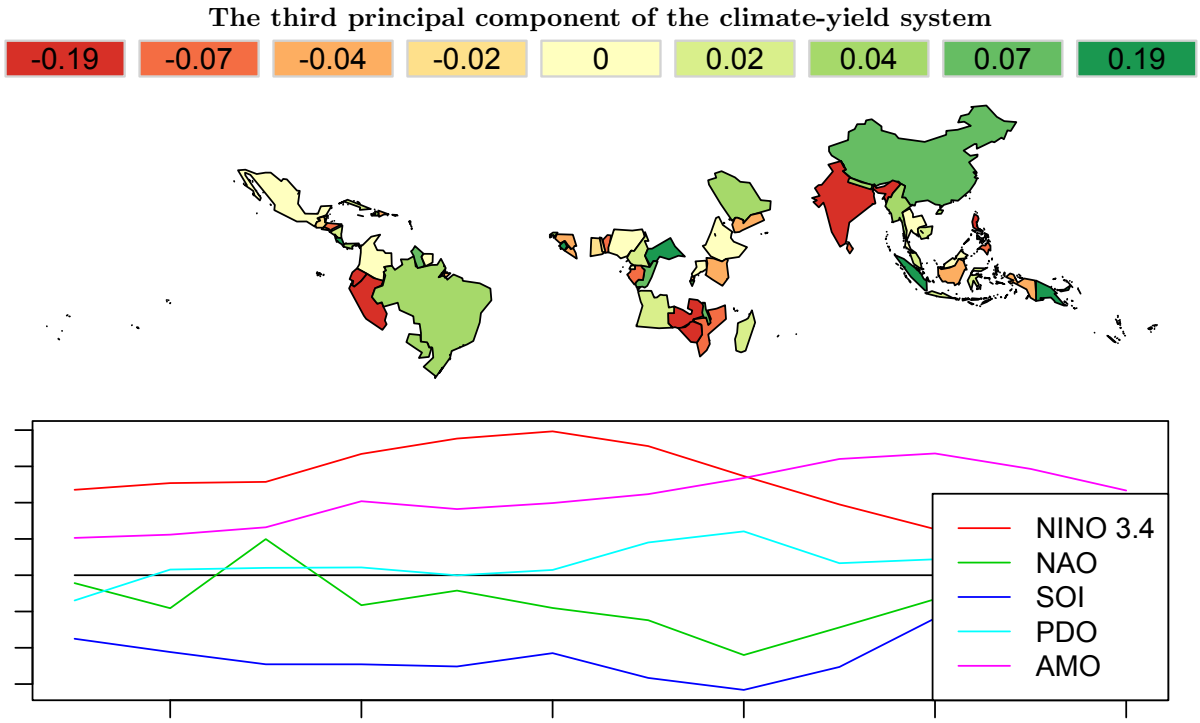


Figure 37: Spatial and temporal representation of the third principal component of the climate-yield system. Colors in the map represent increases (green) and decreases (red), and the plot below shows the climate signals across the year (delayed 6 months). Explanation in the text.

The third principal component also occurs when ENSO is in its El Niño state, and AMO is high or increasing. In this case, India, Peru, and southern areas in Africa show decreases, while other areas are not heavily affected. Both PC 2 and PC 3 can equally be understood in their La Niña form (and associated low values of PDO for PC 2 and low values of AMO for PC 3), which produce changes in yields in the opposite direction.

Between PCs 2 and 3, the effects of El Niño and La Niña appear across much of the globe. Because the impacts on most countries result from an interaction between the ENSO cycle and AMO or PDO, the results did not appear in the initial analysis.

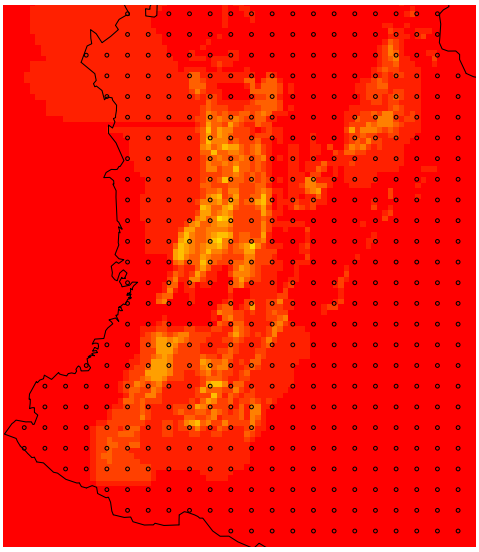
0.7.5 Spatially-weighting weather

To generate weather observations at the same spatial aggregation as yields, we perform the following procedure. For each political unit,

1. Translate CFSR grid cells into a lattice of points.

2. Find all grid lattice points within a given country.
3. Identify the measure of harvested area in the coffee database nearest to each lattice point.
4. Take the weighted average of weather observations, weighted by coffee harvested area.

An example is shown below for grid cells that fall within Colombia.



Circles show the location of CFSR grid lattice points. Colors show the coffee weights.

0.8 Extra market analysis

0.8.1 Market data

The coffee market model incorporates coffee production divided out by producer countries, coffee consumption divided out by consuming countries, and the national and international drivers of the prices paid to growers and by consumers. The following inputs are used to construct an empirically-grounded market model. All are available at least at a yearly resolution, and are here implicitly indexed by year.

q_i	Production in country i	USDA and FAO
p_i	Price to growers in country i	ICO
d_j	Consumption in country j	UN Comtrade
c_j	Retail price in country j	ICO
e_{ij}	Export from country i to j	UN Comtrade
u	International coffee price	World Bank

Below, we use a consumer price index[4] to translate prices to year 2000 US dollars, as shown for Arabica and Robusta international prices in figure 38. Coffee consumers have enjoyed a significant reduction in prices, in real terms, since the 1970s and 1980s.

[4]We use a single CPI across all countries, calculated by International Financial Statistics for their "All Items" goods in advanced economies.

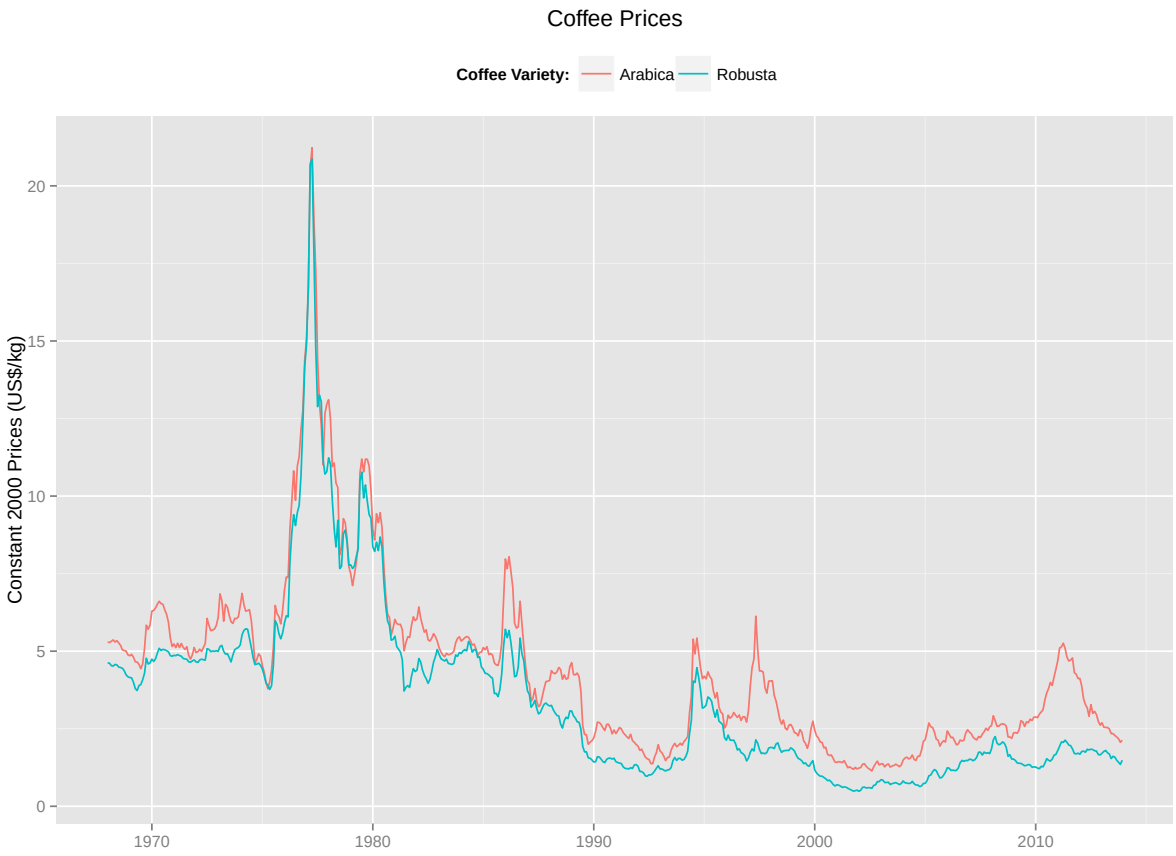


Figure 38: Arabica and Robusta green bean coffee prices, in terms of constant year 2000 US$ per kilogram.

The following sections estimate empirical price relationships. While these are greatly simplified, they provide approximation to the drivers of the international coffee market.

0.8.2 International prices and production

International prices a partially driven by global production, but with considerable autocorrelation. The simplest form of this relationship is:

$$u = \alpha_0 + \alpha_1 \sum_i q_i + \alpha_2 u_{t-1}$$

This expression represents the fundamental driver of international coffee prices: scarcity increases prices and a glut of coffee reduces them. α_1 is negative to capture this relationship. α_2 represents the extent to which prices adjust slowly and are driven by other shocks. If α_2 is near 1, coffee prices have a long memory; while if it is near 0, they respond immediately to production changes. The result is estimated in table 13.

	Dependent variable:	
	Arabica	Robusta
	(1)	(2)
α_0	0.059	0.074
	(0.075)	(0.060)
α_1	−0.00000	−0.00001
	(0.00001)	(0.00002)
α_2	0.988***	0.987***
	(0.007)	(0.007)
Observations	551	551
R^2	0.975	0.980
Adjusted R^2	0.975	0.980
Residual Std. Error (df = 548)	0.458	0.410
F Statistic (df = 2; 548)	10,669.730***	13,612.970***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 13: Estimate of the effect of production on international prices.

This estimate places the entire weight of the predictive capacity on the autoregressive term. In other words, the only significant information about the future international price is the current international price.

To improve this analysis, we model the dynamics of Arabica and Robusta stocks as a closer proxy for the driving quantities on the market, in Appendix 0.8.6. We find that international prices continue to be best explained by their own internal dynamics.

0.8.3 Prices to growers

Prices to growers are affected by both international prices and local production:

$$p_i = \beta_0 + \beta_1 q_i + \beta_2 u$$

Farmers are paid more when coffee fetches a higher price on the international market ($\beta_2 > 0$) , but less if there is a relative excess of coffee produced in their country in a given year ($\beta_1 < 0$). We further allow for country-specific unexplained variation. The results of this estimate are shown in figure 4.5 and in the table in Appendix 0.8.7. The data is from International Coffee Organization (2015).

0.8.4 Consumer response to prices

We expect consumption to decrease with retail prices:

$$d_j = \gamma_0 + \gamma_1 c_j + \gamma_2 d_{j,t-1}$$

Demand does not re-calibrate immediately to changes in retail prices ($\gamma_2 > 0$), but we assume that high prices produce a downward force while low prices produce an upward force ($\gamma_1 < 0$). However, we allow for this "economic" force to be dominated by internal consumption dynamics, represented here as external demand shocks that persist through the autoregression term γ_2. These results are shown in figure 4.9 and in table 16 in the Appendix. The data is from International Coffee Organization (2015).

0.8.5 Retail prices follow costs

Retail costs are a composite of the costs for imports from each country, plus a markup:

$$c_j = \phi_j + \sum_i \frac{e_{ij}}{d_j} (p_i + \theta_i + l_{ij})$$

Retailers respond to the costs of their inputs, which combine country-specific production prices (p_i), added prices for processing and tariffs (θ_i), a cost related to the transportation between them (l_{ij}), and added costs specific to the retailing country (ϕ_j). The extent to which each of the producing country variables (p_i, θ_i, l_{ij}) impact the final retail price is determined by the faction imported from each country (e_{ij}/d_j).

0.8.6 Stock analysis

As an improvement, we note that prices are determined more directly by the stocks of coffee beans available to coffee markets. Therefore, we explore adding stocks of Arabica and Robusta to the model. These are not recorded separately by variety, although the USDA Foreign Agricultural Service reports total coffee stocks. We use a Bayesian model to infer the stocks, informed simultaneously by monthly production and the ability of these stocks to inform prices. The inferred stocks are shown in figure 39.

Reported stocks were much higher than those inferred by the model. This is because the model attempts to use low initial stocks to explain the high international coffee prices in the 1970s. Later, bursts in stock correspond closely with increases in recorded stocks– for example, in 1982, 1988, and 2003. However,

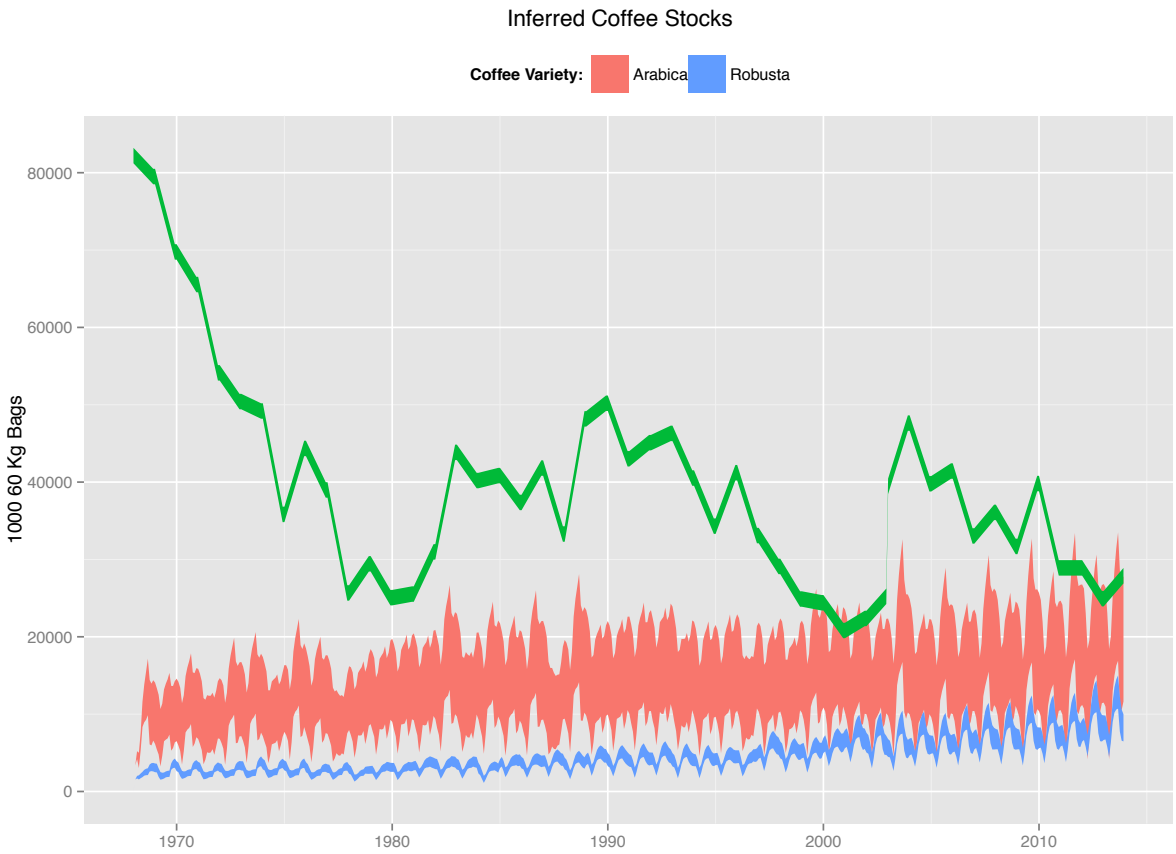


Figure 39: Inferred stocks of Arabica (red) and Robusta (blue) coffee, compared with reported stocks (green) summed over all countries. Arabica and Robusta curves are shown with 50% confidence intervals, while the recorded curve is shown with a constant width.

the model predicts a rapid decrease in the stock after the burst, while recorded stocks remain high after each event. This could reflect the sensitivity of the market to "fresh" green beans, rather than stored ones.

The stocks are estimated simultaneously with the effect they have on the prices, thereby using the price to inform the level of stock. We further estimate the price in logs, and add an effect of the CPI (c_t), to produce our final model:

$$\log u = \alpha_0 + \alpha_1 s_i + \alpha_2 \log u_{t-1} \alpha_3 \log c_t$$

The estimates and their standard deviations are shown in table 14. The effect of stock levels is still negative as expected, but not statistically significant. The autocorrelation in α_2 is decreased because of the other informative elements. The coefficients also suggest that as CPI increases, international prices decrease, although this might just reflect the general downward trend in prices.

	Arabica		Robusta	
	Mean	Std. Dev.	Mean	Std. Dev.
α_0	0.96	1.08	0.48	0.6
α_1	-8.47×10^{-7}	4.32×10^{-6}	-8.85×10^{-6}	1.17×10^{-5}
α_2	0.73	0.31	0.89	0.14
α_3	−0.16	0.17	−0.08	0.1

Table 14: Coefficient estimates for the full stock model. Only α_2, the coefficient on delayed international price, is significant at a 95% level.

0.8.7 Explaining prices to farmers

Estimates for the determinants of prices to growers of coffee.

Country	Type	International	Log Product.	Int. V.E.	Prod. V.E.	Int. P(>|t|)	Prod. P(>|t|)
Colombia	Colombian Milds	16.190	-0.004	0.855	0.033	0.000	0.001
Kenya	Colombian Milds	42.872	-0.037	0.958	0.013	0.000	0.001
Tanzania	Colombian Milds	26.367	0.110	0.856	0.022	0.000	0.022
Bolivia	Other Milds	26.126	0.126	0.831	0.002	0.000	0.635
Burundi	Other Milds	13.918	0.006	0.768	0.000	0.000	0.819
Cameroon	Other Milds	17.297	-0.043	0.509	0.021	0.000	0.351
Congo, Dem. Rep. of	Other Milds	5.788	-0.148	0.275	0.708	0.297	0.098
Costa Rica	Other Milds	21.069	-0.052	0.839	0.066	0.000	0.000
Cuba	Other Milds	14.087	0.584	0.337	0.455	0.000	0.000
Dominican Republic	Other Milds	32.377	-0.040	0.937	0.006	0.000	0.046
Ecuador	Other Milds	23.186	-0.025	0.807	0.010	0.000	0.151
El Salvador	Other Milds	33.379	-0.018	0.930	0.011	0.000	0.010
Guatemala	Other Milds	20.445	-0.019	0.881	0.025	0.000	0.002
Haiti	Other Milds	19.746	0.097	0.888	0.012	0.000	0.094
Honduras	Other Milds	35.087	0.009	0.910	0.007	0.000	0.072
India	Other Milds	17.230	-0.039	0.873	0.023	0.000	0.006
Jamaica	Other Milds	18.733	1.791	0.428	0.038	0.000	0.113
Madagascar	Other Milds	0.211	-0.074	0.000	0.001	0.987	0.933
Malawi	Other Milds	28.364	0.277	0.731	0.102	0.000	0.004
Mexico	Other Milds	25.079	-0.023	0.892	0.018	0.000	0.016
Nicaragua	Other Milds	39.074	-0.010	0.877	0.001	0.000	0.641
Panama	Other Milds	13.132	-0.131	0.665	0.008	0.000	0.451
Papua New Guinea	Other Milds	21.258	-0.023	0.888	0.003	0.000	0.302
Peru	Other Milds	47.729	-0.013	0.778	0.006	0.000	0.458
Rwanda	Other Milds	13.679	0.084	0.706	0.069	0.000	0.005
Uganda	Other Milds	2.894	0.043	0.040	0.083	0.039	0.053
Venezuela	Other Milds	22.230	0.138	0.650	0.048	0.001	0.261
Zambia	Other Milds	13.538	-0.041	0.559	0.000	0.000	0.852
Zimbabwe	Other Milds	14.874	3.177	0.681	0.055	0.455	0.585
Brazil	Brazilian Naturals	22.184	0.001	0.794	0.016	0.000	0.067
Ethiopia	Brazilian Naturals	14.567	-0.006	0.765	0.015	0.000	0.097
Indonesia	Brazilian Naturals	31.851	0.005	0.528	0.001	0.011	0.919
Philippines	Brazilian Naturals	28.897	-0.040	0.414	0.001	0.005	0.901
Angola	Robustas	6.650	0.010	0.232	0.033	0.019	0.241
Benin	Robustas	0.449	-0.637	0.155	0.252	0.711	0.016
Brazil	Robustas	27.762	0.002	0.583	0.077	0.000	0.028
Burundi	Robustas	5.382	1.301	0.082	0.696	0.210	0.055
Cameroon	Robustas	10.715	0.012	0.620	0.012	0.000	0.283
Central African Republic	Robustas	7.394	0.039	0.602	0.018	0.000	0.196

Congo, Dem. Rep. of	Robustas	10.625	-0.023	0.610	0.055	0.000	0.059
Congo, Rep. of	Robustas	1.318	0.227	0.333	0.137	0.047	0.030
Côte d'Ivoire	Robustas	9.459	0.006	0.667	0.036	0.000	0.030
Ecuador	Robustas	36.988	-0.022	0.764	0.018	0.000	0.154
Gabon	Robustas	9.228	0.409	0.514	0.018	0.000	0.286
Ghana	Robustas	6.622	-0.721	0.663	0.164	0.023	0.123
Guinea	Robustas	14.037	-0.434	0.537	0.045	0.000	0.157
India	Robustas	23.426	0.007	0.743	0.043	0.000	0.030
Indonesia	Robustas	25.263	0.003	0.942	0.002	0.000	0.339
Madagascar	Robustas	6.512	0.004	0.347	0.001	0.001	0.813
Nigeria	Robustas	8.682	-0.439	0.525	0.034	0.001	0.240
Papua New Guinea	Robustas	28.154	0.028	0.783	0.000	0.000	0.923
Philippines	Robustas	23.835	-0.007	0.913	0.000	0.000	0.780
Sierra Leone	Robustas	11.475	0.181	0.563	0.034	0.000	0.245
Sri Lanka	Robustas	31.224	-0.183	0.930	0.005	0.001	0.547
Tanzania	Robustas	16.526	-0.018	0.585	0.003	0.000	0.689
Thailand	Robustas	28.468	-0.022	0.906	0.004	0.000	0.193
Togo	Robustas	5.642	0.022	0.472	0.016	0.000	0.255
Trinidad & Tobago	Robustas	24.719	0.426	0.555	0.001	0.002	0.786
Uganda	Robustas	2.255	-0.005	0.153	0.019	0.026	0.354
Vietnam	Robustas	77.812	-0.001	0.576	0.002	0.000	0.745
Global	All	25.651	-0.002	0.463	0.000	0.000	0.000

0.8.8 Explaining consumer demand

Country	Retail Pr.	Previous Yr.	Ret. V.E.	Prev. V.E.	Ret. Pr(>|t|)	Prev. Pr(>|t|)
Austria	0.01	0.35	0.00	0.13	0.88	0.14
Belgium	-0.47	1.13	0.30	0.35	0.00	0.01
Bulgaria	0.15	0.20	0.36	0.04	0.03	0.37
Cyprus	-0.02	0.23	0.07	0.05	0.24	0.28
Czech Republic	-0.01	0.69	0.00	0.32	0.37	0.01
Denmark	0.03	0.83	0.10	0.61	0.53	0.00
Finland	0.04	0.03	0.01	0.00	0.70	0.88
France	-0.00	0.66	0.02	0.42	0.95	0.00
Germany	-0.08	-0.35	0.17	0.08	0.04	0.22
Hungary	-0.06	0.86	0.02	0.68	0.07	0.00
Italy	0.02	0.79	0.11	0.49	0.74	0.00
Latvia	0.01	0.72	0.16	0.67	0.32	0.00
Lithuania	0.01	0.83	0.14	0.78	0.18	0.00
Luxembourg	-1.42	0.52	0.53	0.16	0.06	0.03
Malta	0.00	0.43	0.01	0.13	0.68	0.15
Netherlands	-0.05	0.74	0.00	0.60	0.76	0.00
Poland	0.00	0.77	0.00	0.49	0.86	0.00
Portugal	-0.02	0.79	0.14	0.67	0.14	0.00
Slovakia	0.00	-0.46	0.00	0.18	0.93	0.08
Slovenia	0.02	0.21	0.38	0.04	0.03	0.32
Spain	0.01	0.91	0.51	0.33	0.72	0.00
Sweden	-0.04	0.30	0.08	0.10	0.18	0.15
United Kingdom	0.01	0.36	0.01	0.11	0.80	0.35
Japan	0.00	0.78	0.56	0.17	0.64	0.00
Norway	0.03	0.23	0.03	0.05	0.58	0.34
Switzerland	-0.04	0.58	0.44	0.12	0.27	0.02
Turkey	-0.00	1.02	0.16	0.50	0.82	0.00
USA	-0.01	0.12	0.07	0.01	0.35	0.64
Global	-0.01	0.70	0.01	0.92	0.08	0.00

Table 16: Determinants of consumption of coffee.

0.8.9 Inferred markups

Producer country markups

Mark-ups over the prices paid to farmers, by producer country, in US cents per pound. These are estimated simultaneous with the consumer country mark-ups.

Country	To Farmers	Mark Up	Std. Dev.
Bolivia	148.59	223.80	112.91
Brazil	107.06	84.80	31.14
Burundi	91.43	242.62	113.93
Cameroon	79.02	239.24	105.61
Sri Lanka	59.24	228.88	117.07
Colombia	124.50	194.32	70.11

Country	To Farmers	Mark Up	Std. Dev.
Congo, Dem. Rep. of	52.37	257.35	116.00
Costa Rica	137.20	214.23	102.04
Cuba	212.77	256.45	122.19
Dominican Republic	135.59	236.17	113.14
Ecuador	110.42	246.34	115.02
El Salvador	113.71	204.70	99.97
Ethiopia	101.05	168.01	94.98
Gabon	84.70	228.00	113.03
Ghana	108.59	244.26	121.38
Guatemala	137.95	102.30	70.70
Guinea	126.11	212.57	109.47
Haiti	91.81	229.51	114.36
Honduras	126.66	292.54	113.00
Indonesia	97.53	339.22	88.09
Côte d'Ivoire	63.74	226.83	96.66
Jamaica	262.39	239.55	116.14
Kenya	198.15	239.01	109.90
Madagascar	72.44	218.01	109.38
Malawi	77.87	235.31	117.05
Mexico	155.06	221.07	106.49
Nicaragua	139.47	182.82	96.48
Panama	149.15	235.36	106.61
Papua New Guinea	102.78	177.36	100.33
Peru	123.27	243.61	106.00
Rwanda	95.20	230.81	113.27
India	114.96	196.25	92.03
Vietnam	120.97	29.68	25.36
Thailand	93.03	268.90	122.50
Togo	55.70	225.31	108.66
Uganda	44.89	246.98	80.22
Tanzania	111.59	254.27	120.02
Venezuela	258.16	230.70	115.24
Zambia	110.58	219.19	112.56
Congo, Rep. of	33.52	226.88	120.91
Nigeria	118.40	228.59	114.78
Sierra Leone	111.94	234.04	110.40
Zimbabwe	408.97	234.58	118.08
Central African Republic	58.13	229.93	116.94
Trinidad & Tobago	173.12	228.67	114.16
Philippines	107.48	241.64	119.16
Angola	77.44	255.29	117.56
Benin	63.19	233.40	112.44
Liberia	NA	234.01	115.80
(Processed)	NA	254.26	34.22

Consumer country markups

Country	To Farmers	Distribution	Retail	Mark Up	Std. Dev.
Austria	89.40	184.31	448.67	175.07	34.15
Belgium	79.54	173.86	392.38	139.93	38.09
Bulgaria	74.88	208.52	283.88	31.51	25.96
Cyprus	86.47	112.54	429.70	230.73	35.58
Denmark	84.43	179.78	447.29	184.20	33.37
Finland	93.54	162.29	309.97	58.43	31.33
France	82.91	176.00	283.45	37.98	26.30
Germany	83.16	136.86	380.83	162.09	35.53
Hungary	68.32	205.45	387.22	114.71	39.63
Italy	81.08	132.38	571.66	356.53	37.01
Latvia	82.82	249.52	441.37	109.76	42.82
Lithuania	86.29	248.47	422.60	89.78	43.23
Luxembourg	87.45	254.18	559.21	217.24	48.44
Malta*	92.70	228.81	1019.96	692.57	41.21
Netherlands	88.05	192.84	366.49	86.88	34.92
Poland	71.99	178.79	317.54	70.10	32.04
Portugal	74.36	202.26	484.34	206.79	34.62
Slovakia	76.42	219.32	342.90	54.81	31.93
Slovenia	77.50	144.39	367.10	146.78	38.11
Spain	77.46	165.04	350.89	109.90	31.78
Sweden	93.92	168.28	350.25	89.53	34.55
United Kingdom*	81.04	190.61	1354.07	1070.48	37.39
Japan	91.23	182.71	1107.51	828.59	35.21
Norway	98.18	152.83	372.36	122.48	36.67
Switzerland	84.83	184.49	524.88	254.33	35.19
Turkey	91.25	99.09	416.92	226.54	39.70
USA	85.46	133.48	345.80	127.57	37.40

Table 18: Mark-ups over the prices paid to farmers, by consumer country, in US cents per pound. These are estimated simultaneous with the producer country mark-ups.

0.8.10 Economic importance of coffee

Country	Variety	Farm ($/kg)	Itnl. ($/kg)	Production (kg)	Local Value ($)	Intl Value ($)	GDP ($)	(%)
Angola	arabica	0.82	2.74	0	0	0	44080659100	0.00
Angola	robusta	0.59	1.46	1698000	1004067	2483891	44080659100	0.01
Bolivia	arabica	3.70	2.74	8442000	31232762	23162986	11362240323	0.20
Brazil	arabica	2.12	2.74	2026500000	4294893862	5560268994	1019917358692	0.55
Brazil	robusta	1.39	1.46	711900000	991280609	1041391089	1019917358692	0.10
Burundi	arabica	1.28	2.74	16026000	20465292	43971809	1320071226	3.33
Burundi	robusta		1.46	0		0	1320071226	0.00
Cameroon	arabica	1.41	2.74	5532000	7807333	15178588	18629569264	0.08
Cameroon	robusta	0.97	1.46	39630000	38314825	57972087	18629569264	0.31
Central African Republic	robusta	0.91	1.46	1518000	1378446	2220581	1631792478	0.14
Colombia	arabica	2.38	2.74	597966000	1422490591	1640686804	175186920716	0.94
Congo, Dem. Rep. of	arabica		2.74	5262000		14437767	14788569483	0.10
Congo, Dem. Rep. of	robusta		1.46	11724000		17150259	14788569483	0.12
Congo, Rep. of	robusta		1.46	0		0	7163015187	0.00
Costa Rica	arabica	2.15	2.74	104058000	223638738	285512199	23899924562	1.19
Côte d'Ivoire	robusta	0.61	1.46	121800000	73750900	178173107	18513289732	0.96
Cuba	arabica	1.39	2.74	7740000	10766855	21236853	49991611345	0.04
Dominican Republic	arabica	2.14	2.74	27162000	58194207	74526537	42066031892	0.18
Ecuador	arabica	2.45	2.74	25110000	61503896	68896301	47883528950	0.14
Ecuador	robusta	1.30	1.46	17400000	22547442	25453301	47883528950	0.05
El Salvador	arabica	1.70	2.74	87264000	148428732	239433167	18219072282	1.31
Ethiopia	arabica	1.47	2.74	315408000	465181914	865409979	18472707950	4.68
Ghana	robusta		1.46	2214000		3238713	14161252979	0.02
Guatemala	arabica	2.24	2.74	232476000	521865895	637862864	31344823589	2.03
Guinea	robusta		1.46	23412000		34247855	3208509564	1.07
Haiti	arabica		2.74	19368000		53141520	4429227578	1.20
Honduras	arabica	1.92	2.74	227994000	438103331	625565245	11134160781	5.62
India	arabica	2.35	2.74	98028000	230277245	268967209	1109531099998	0.02
India	robusta	1.43	1.46	187410000	268898827	274149605	1109531099998	0.02
Indonesia	arabica	2.09	2.74	69150000	144393359	189732347	353469522531	0.05
Indonesia	robusta	0.82	1.46	461700000	377800422	675390175	353469522531	0.19
Jamaica	arabica	5.55	2.74	1686000	9360707	4626012	11075778481	0.04
Kenya	arabica	1.45	2.74	47946000	69690082	131553248	22500842427	0.58
Liberia	robusta		1.46	1836000		2685762	839609172	0.32
Madagascar	arabica	0.99	2.74	1632000	1611367	4477848	5652013484	0.08
Madagascar	robusta	0.79	1.46	30264000	23879354	44271190	5652013484	0.78
Malawi	arabica	1.54	2.74	1560000	2402418	4280296	3465670842	0.12
Mexico	arabica	2.00	2.74	251100000	502945556	688963012	94313175944	0.07
Nicaragua	arabica	1.17	2.74	99240000	115978241	272292670	7096139115	3.84
Nigeria	robusta		1.46	2226000		3256267	144037995041	0.00

Country	Variety	Farm ($/kg)	Itnl. ($/kg)	Production (kg)	Local Value ($)	Intl Value ($)	GDP ($)	(%)
Panama	arabica		2.74	7044000		19327182	2125364815	0.09
Papua New Guinea	arabica	1.59	2.74	59610000	94843298	163556691	6141247952	2.66
Papua New Guinea	robusta	0.56	1.46	2748000	1542012	4019866	6141247952	0.07
Peru	arabica	1.13	2.74	224640000	253661686	616362609	96436263271	0.64
Philippines	arabica	2.01	2.74	1950000	3915230	5350370	123566704265	0.00
Philippines	robusta	1.33	1.46	27990000	37124998	40944707	123566704265	0.03
Rwanda	arabica	0.76	2.74	17640000	13479953	48400269	3504265695	1.38
Sierra Leone	robusta	0.44	1.46	3414000	1517699	4994113	1998841063	0.25
Sri Lanka	arabica		2.74	594000		1629805	31081582880	0.01
Sri Lanka	robusta		1.46	1446000		2115257	31081582880	0.01
Tanzania	arabica	0.91	2.74	31134000	28441963	85424828	18155123928	0.47
Tanzania	robusta	0.38	1.46	19248000	7401039	28156617	18155123928	0.16
Thailand	arabica	3.50	2.74	0	0	0	199643270642	0.00
Thailand	robusta	1.30	1.46	53166000	69270076	77773000	199643270642	0.04
Togo	robusta	0.96	1.46	20520000	19786230	30017341	2406177565	1.25
Trinidad & Tobago	robusta		1.46	336000		491512	18345375584	0.00
Uganda	arabica	1.37	2.74	37038000	50823541	101624102	12065197411	0.84
Uganda	robusta	1.03	1.46	143748000	148064470	210279374	12065197411	1.74
Venezuela	arabica		2.74	50298000		138006617	171452404755	0.08
Vietnam	arabica	1.47	2.74	30684000	45130389	84190128	72353243154	0.12
Vietnam	robusta	1.26	1.46	1108206000	1395836280	1621120737	72353243154	2.24
Zambia	arabica	1.91	2.74	3132000	5972119	8593517	11256721495	0.08
Zimbabwe	arabica		2.74	2538000		6963712	5603223515	0.12

Coffee production database

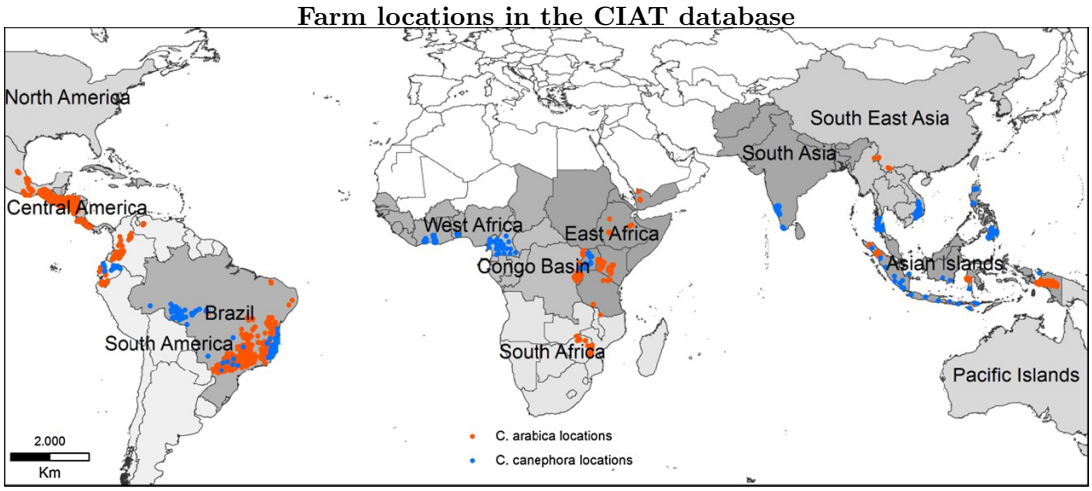


Figure 40: Arabica and Robusta coffee producing farms in the CIAT database.

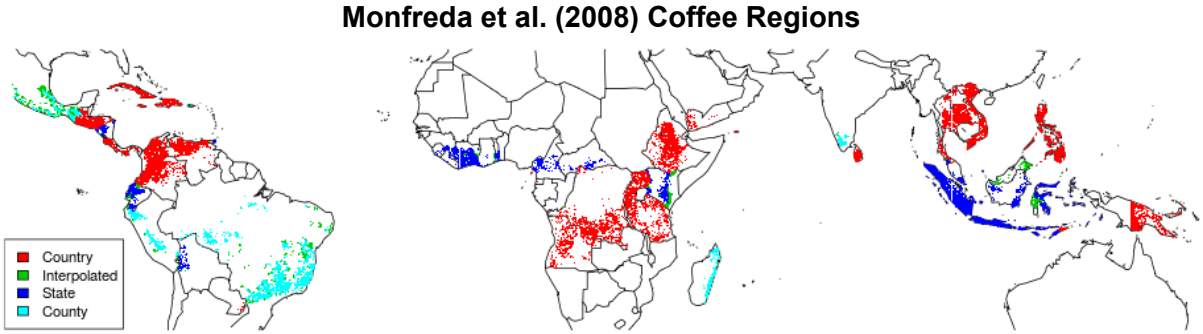


Figure 41: Quality of geospatial production data for coffee, from Monfreda et al. (2008). Global agricultural areas are intersected with country-specific datasets to create a global map of coffee production areas.

0.8.11 Confidence maps

The first result of the database is its own measure of confidence in the geographic data across the globe. The confidence maps reflect the combined amount of information available, across the multiple map inputs. Each contributing map is assigned its own confidence, with maps of global harvest having low or medium confidence and maps detailing a given country with high confidence. Where multiple input maps corroborate each other, the confidence increases (see appendix 0.10.1). In figure 42, dark green represents low confidence, and yellow and tan colors represent high confidence. The band of lighter green in the middle shows the overlap between maps from Thurston et al. (2013), Monfreda et al. (2008), and Bunn et al. (2015).

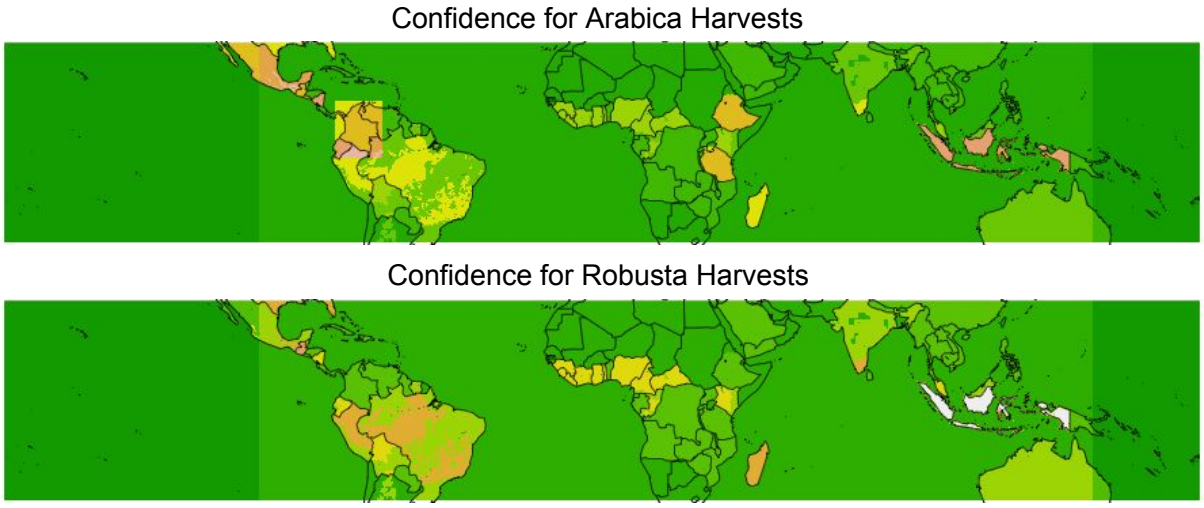


Figure 42: Database geospatial harvest confidence, based on the amount and scale of data available.

0.8.12 Harvest maps

The harvest maps are the main output of the spatial portion of coffee database. For each month, these combine country-specific information (some of which specifies harvest months as they differ across the country), with global harvesting regions applied to a calendar of harvest months from Sweet Maria (2015). Some country calendars are unavailable, so these harvested regions show throughout the year. The added weight of these multiple instances will be handled next.

0.8.13 Time series data

Both FAO and the USDA Foreign Agricultural Service report production information for coffee, but the information they provide is quite different. FAO reports total production and harvested area, for all varieties of coffee combined, with a total of 4242 observations. The USDA reports only production information, but divides it out by Arabica and Robusta production, with 3211 observations per variety. The number of countries included also varies by year (see figure 44).

A second complication arises from the definition of the reported year. FAO reports production for calendar years, while USDA reports it for market years which vary by country. This can be an opportunity,

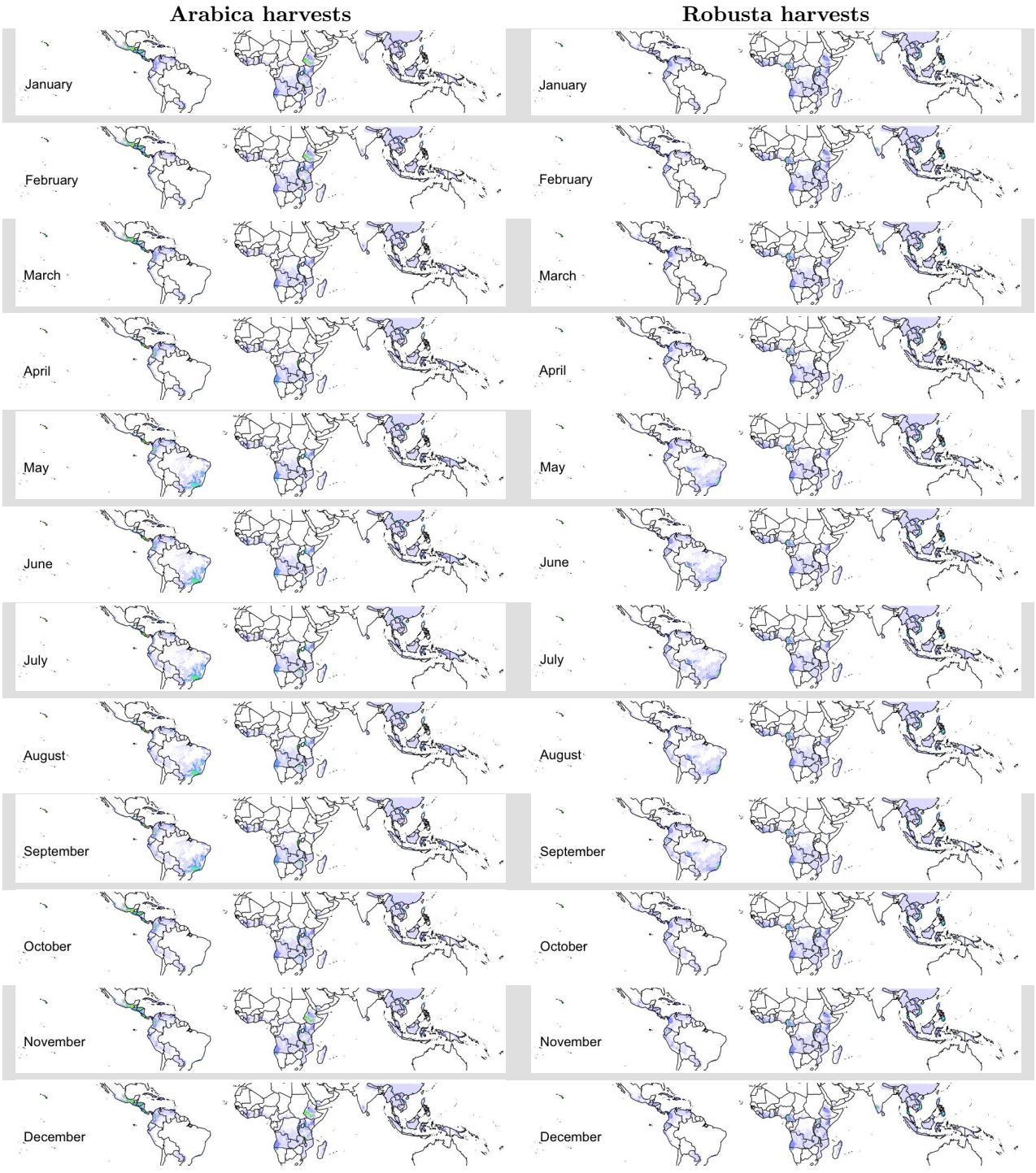


Figure 43: Harvest maps for Arabica and Robusta varieties, during each month. Darker colors represent higher levels of evidence that these regions are undergoing harvest in the given month.

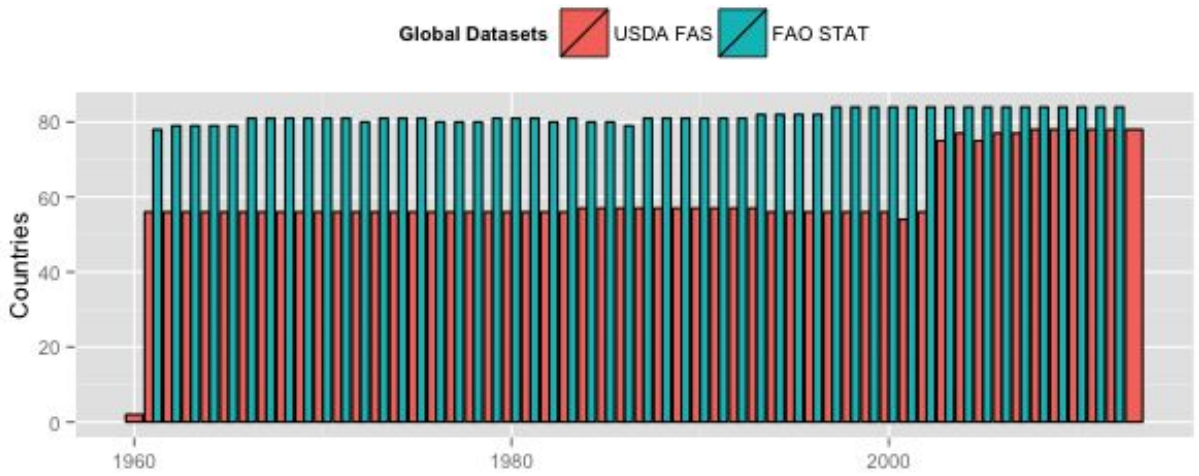
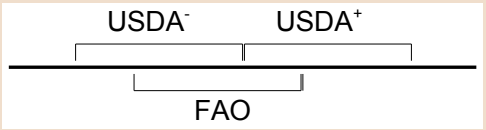


Figure 44: The number of countries with coffee production data available by year.

allowing us to determine more precisely when production occurs. For example, in Brazil, coffee is harvested mainly between May and September. However, the USDA market year for Brazil is from July to June. So, discrepancies between the FAO and USDA production totals allow us to distinguish, approximately, between the share of production before and after July, the start of the market year cycle.

Calculating intra-year production

The diagram below shows how the USDA and FAO calendars align. The actual division is different for each country, depending on the start of the USDA market year.



We divide each USDA value into "left" and "right" parts, with $USDA^L = \alpha USDA$ and $USDA^R = (1-\alpha)USDA$, where the coefficient α is unknown. Further, we know from the diagram that $USDA^R_- + USDA^L_+ = FAO$; that is, the FAO year consists of the 'right' (latter) portion of one market year and the 'left' (early) portion of the next one. Finally, we can use the difference to estimate α from

$$FAO_t - USDA_{t+} = \alpha(USDA_{t-} - USDA_{t+}) + \epsilon_t$$

We estimate this division for each country. In the case of Brazil, we find that 12% of production occurs between May and June, and 88% between July and September. A full table of these portions is shown in table 20. Where there are blanks, the two datasets could not be consistently combined.

Using these values, we can construct a monthly timeseries of production, as shown in figure 45.

The coffee database consists of paired production and growing region files. The database consists of both the final files and the code for generating standardized versions of input source files. The standardized versions have the same format as the merged database.

The coffee database is available in a sharable form, at `https://bitbucket.org/jrising/coffeedb/`. Request for access.

	Market Year	Previous Year	Following Year	Std. Err.
Brazil	Jul - Jun	0.12	0.88	0.05
Madagascar	Apr - Mar	0.50	0.50	0.25
Kenya	Oct - Sep	0.91	0.09	0.04
Guinea	Oct - Sep	0.37	0.63	0.33
Panama	Oct - Sep	0.68	0.32	0.28
Costa Rica	Oct - Sep	0.50	0.50	0.06
Ethiopia	Oct - Sep			
Rwanda	Apr - Mar	0.17	0.83	0.08
United Republic of Tanzania	Jul - Jun	0.67	0.33	0.11
Sri Lanka	Oct - Sep			
Peru	Apr - Mar	0.07	0.93	0.10
Lao People's Democratic Republic	Oct - Sep			
Bolivia (Plurinational State of)	Apr - Mar			
Cameroon	Oct - Sep	0.09	0.91	0.11
Côte d'Ivoire	Oct - Sep	0.89	0.11	0.07
Ecuador	Apr - Mar	0.41	0.59	0.25
Benin	Oct - Sep	0.60	0.40	0.20
Ghana	Oct - Sep	0.80	0.20	0.16
Cuba	Jul - Jun	0.46	0.54	0.16
El Salvador	Oct - Sep	0.40	0.60	0.05
Venezuela (Bolivarian Republic of)	Oct - Sep	0.57	0.43	0.15
Papua New Guinea	Apr - Mar	0.26	0.74	0.07
Malawi	Oct - Sep	0.10	0.90	0.16
Togo	Oct - Sep	0.44	0.56	0.16
Guatemala	Oct - Sep	0.42	0.58	0.17
Zimbabwe	Oct - Sep	0.19	0.81	0.10
Viet Nam	Oct - Sep	0.63	0.37	0.07
Dominican Republic	Jul - Jun	0.55	0.45	0.15
Nigeria	Oct - Sep	0.68	0.32	0.19
Liberia	Oct - Sep	0.56	0.44	0.14
Democratic Republic of the Congo	Oct - Sep	0.61	0.39	0.14
Paraguay	Oct - Sep	0.60	0.40	0.34
Trinidad and Tobago	Oct - Sep	0.77	0.23	0.10
Philippines	Jul - Jun			
Indonesia	Apr - Mar	0.23	0.77	0.29
Central African Republic	Oct - Sep	0.22	0.78	0.14
New Caledonia	Oct - Sep			
United States of America	Oct - Sep	0.19	0.81	0.65
Guyana	Oct - Sep			
Honduras	Oct - Sep	0.56	0.44	0.08
Yemen	Oct - Sep	0.36	0.64	0.86
Haiti	Jul - Jun	0.36	0.64	0.19
Thailand	Oct - Sep	0.77	0.23	0.07
Jamaica	Oct - Sep	0.69	0.31	0.98
Angola	Apr - Mar	0.62	0.38	0.11
Equatorial Guinea	Oct - Sep			
Mexico	Oct - Sep	0.62	0.38	0.22
India	Oct - Sep	0.98	0.02	0.02
Sierra Leone	Oct - Sep	0.98	0.02	0.60
Malaysia	Oct - Sep	0.58	0.42	0.33
Congo	Oct - Sep	0.28	0.72	0.26
Colombia	Oct - Sep	0.72	0.28	0.06
Burundi	Apr - Mar	0.06	0.94	0.09
Gabon	Oct - Sep	0.23	0.77	0.17
Uganda	Oct - Sep	0.83	0.17	0.09
Nicaragua	Oct - Sep	0.13	0.87	0.08
Zambia	Oct - Sep	0.93	0.07	0.35

Table 20: Portion of the production for each market year attributed to the previous calendar year and to the next one.

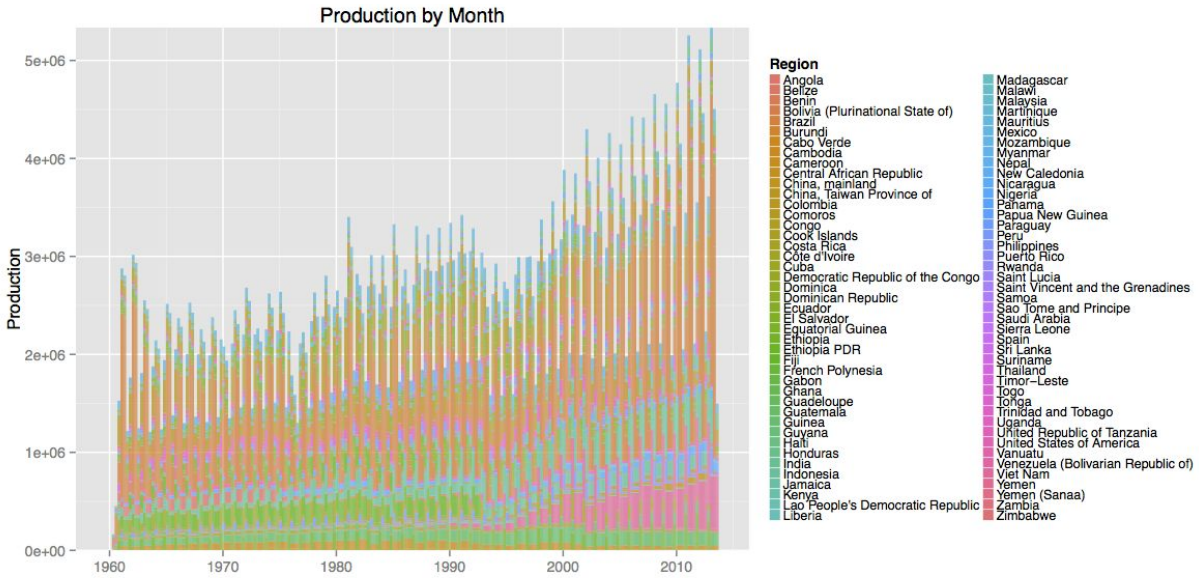
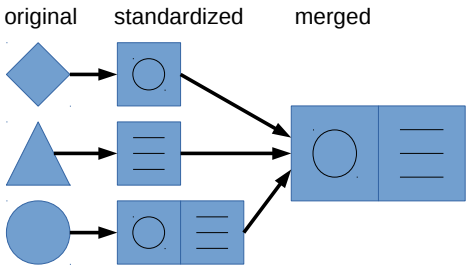


Figure 45: Production by month and country, inferred by the discrepancies between USDA FAS and FAO accounting systems.



0.9 Standardized format

0.9.1 Growing Region Files

Growing regions are stored as raster (gridded) files describe "masks" of which regions are under harvest in a given month. They are at a resolution of 12 pixels per degree (a grid width of 5 minutes), and cover the entire area from 180°W to 180°E longitude, and 30°S to 30°N latitude.

The grids are stored as NetCDF files. Each NetCDF file contains a "harvest" variable and a "confidence" variable. The harvest variable specifies areas under harvest in a given month, and has dimensions Longitude x Latitude x 12, with a separate mask for each month. Values may range between 0 and 1, based on how much evidence there is of harvest there. The confidence mask describes the level of confidence in the information, also from 0 to 1.

Note that neither the "harvest" nor "confidence" variables describe the portion of a given grid cell under harvest. Instead, both relate to the evidence that the grid cell contains areas under harvest. The difference between the harvest value and confidence value is expanded upon in the Merging Growing Region Files section.

0.9.2 Production Files

Production data consists of a `.csv` file that specifies the production, planted area, harvested area, and yields (as data is available for each) in a given year and a given region. Where the data describes sub-country regions, an additional region definition file (`*-regions.csv`) and a shapefile database (collections of a `.shp`, `.shx`, and `.dbf` file) describe an association between the production records and growing regions. Each polygon in the shapefile database identifies a region for which production data is available in one or more years, and region definition files specify which region is described in each record.

The production file has the following column header:

`year,region,variety,produced,prod-se,harvested,harv-se,planted,plant-se,yield,yield-se`

`year` is the year being described. Not all years need to be represented for a region. `variety` is Arabica, Robusta, or combined. `region` is the region identifier in the associated region definitions file. This may change across years. `produced` is the calendar year production, measured in metric tonnes. `prod-se` is the standard error of the production estimate. It may be `NA` if the error estimate is available, but this will cause any other estimate to be chosen over it if one is available. `harvested` is the harvested area in hectares, and `harv-se` is its standard error. This may be `NA`. `planted` is the planted area in hectares, and `plan-se` is its standard error. This may be `NA`. `yield` is the yield in terms of MT per hectare, and `yield-se` is its standard error. The yield is computed as production divided by planted area. This may be `NA`.

The region definitions file has the following column header:

`region,PID,weight`

`region` is a region identifier, unique across the entire database. If there are multiple rows with the same region identifier, all of these PIDs will be combined in the region. `PID` is a polygon IDs in the associated growing region file. The same `PID` may occur in multiple regions, since different regions may be used to describe different years. `weight` is a measure of the accuracy of the production region definitions. In general, `weight` is calculated as *(mean planted area) / (total polygon area)*, and is between 0 (no confidence) and 1 (full confidence).

0.10 Generating merged database files

0.10.1 Merging growing region files

Region definition files are merged according to the weight of evidence of harvest in each month. At every point, the new weight in the combined region definitions is,

$$w(x,y) = \sum_i \frac{w_i(x,y)c_i(x,y)}{c_i(x,y)}$$
$$c(x,y) = \sum_i c_i(x,y)$$

This formulation allows confidence to increase where multiple data sources are available, and causes contradictory maps (for example, one that says that coffee is grown in a region and one that says it is not) to result in averaged values.

	Arabica 1	Arabica 0	Combined 1	Combined 0
Arabica 1, Robusta 0	1, 0	0.5, 0	1, 0	0.5, 0
Arabica 0, Robusta 0	0.5, 0	0, 0	0.5, 0.5	0, 0

Additional logic is used where maps that describe Arabica and Robusta growth separately are combined with those that lump them together.

0.10.2 Merging production files

Production files are merged using a Bayesian approach, with a uniform prior. Each estimate (a given year-region value for production, harvested area, or planted area) is translated into a distribution, $p(y_i)$. Then the merged estimate of production is,

$$p(y) = \prod_i p(y_i)$$

This allows the database to account for uncertainty in the estimates, as well as allowing corroborating records to decrease the amount of uncertainty. Additional logic used where time series that split out Arabica and Robusta growth (such as the USDA Foreign Agricultural Service) are combined with ones that lump them together (such as FAO).

Combining estimates

1. Associate uncertainty with each observation ($k\sqrt{v_{it}}$).
2. Case 1: Same or Combined + (Arabica or Robusta):

$$\mu* = \frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$
$$\sigma*^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

3. Case 2: Combined + Arabica + Robusta:

$$\max_{a,r} \mathcal{N}(a|\mu_a,\sigma_a)\mathcal{N}(r|\mu_r,\sigma_r)\mathcal{N}(a+r|\mu_c,\sigma_c)$$

σ_a* and σ_r* from Inverse Hessian.

0.11 Production maps

We performed a geospatial matching between diagrams in the gray literature and countries maps.

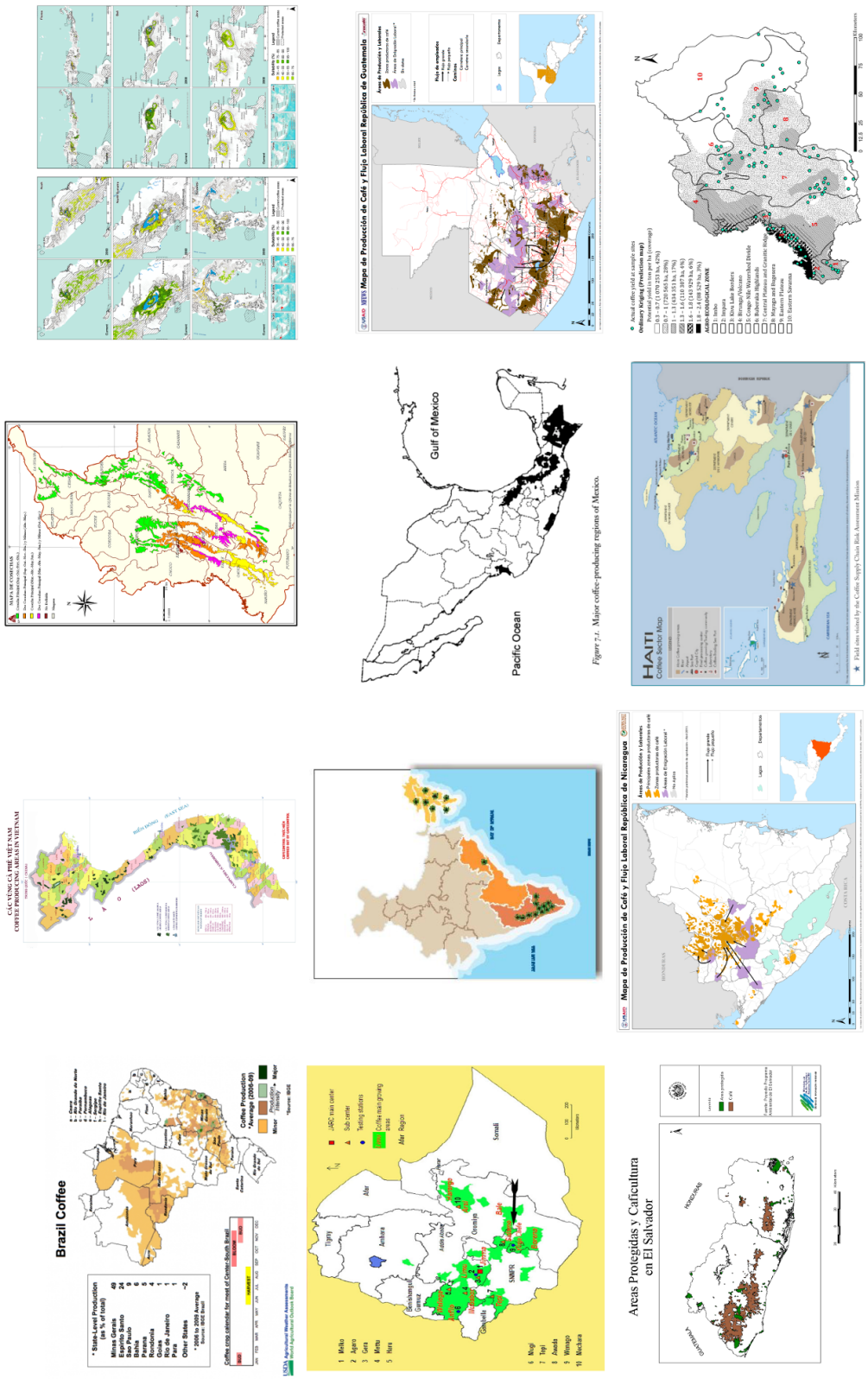


Figure 46: Coffee growing region maps for Brazil, Vietnam, Colombia, Indonesia, Ethiopia, India, Mexico, Guatemala, El Salvador, Nicaragua, Haiti, and Rwanda, collected from various sources (see main text for the table).

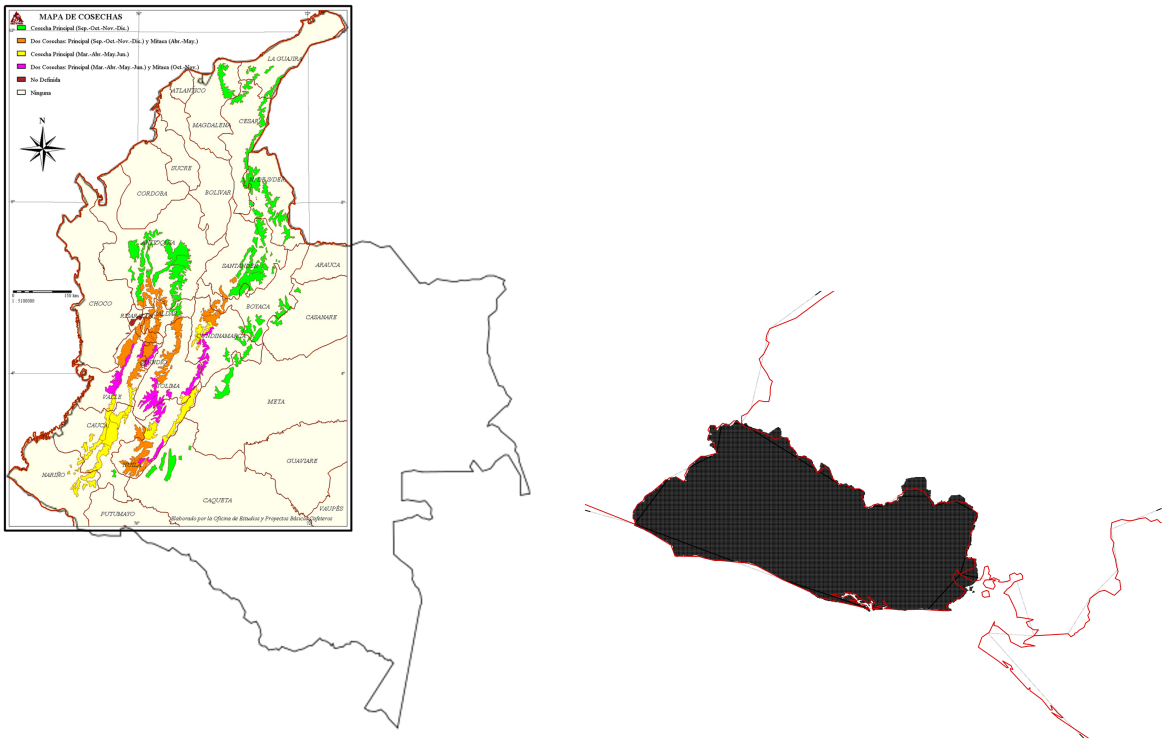


Figure 47: Two examples of the geospatial matching process, using hand correspondences for Colombia (left) and country-wide shape matching for El Salvador (right).

Bibliography

Arbenz, P. (2013). Bayesian copulae distributions, with application to operational risk management—some comments. *Methodology and Computing in Applied Probability*, 15(1):105–108.

Bunn, C., Läderach, P., Rivera, O. O., and Kirschke, D. (2015). A bitter cup: climate change profile of global production of arabica and robusta coffee. *Climatic Change*, 129(1-2):89–101.

Burke, M. and Emerick, K. (2012). Adaptation to climate change: Evidence from us agriculture. *Available at SSRN 2144928*.

Enfield, D. B., Mestas-Nunez, A. M., Trimble, P. J., et al. (2001). The atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental u. s. *Geophysical Research Letters*, 28(10):2077–2080.

FAO/IIASA/ISRIC/ISSCAS/JRC (2012). Harmonized world soil database (hwsd). *FAO, Rome, Italy and IIASA, Laxenburg, Austria*.

Ferreira, S. and Boley, R. (1991). Hemileia vastatrix. *Crop Knowledge Master. Available at: www. extento. hawaii. edu/kbase/crop/type/h_ vasta. htm. Acessed on*, 7(27):2012.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Taylor & Francis.

Guzmán Martínez, O., Jaramillo Robledo, A., and Baldión Rincón, J. V. (1999). Anuario meteorologico cafetero, 1998.

International Coffee Organization (2015). ICO Historical Data. Supplied by ICO, May 7, 2015.

International Research Institute for Climate and Society (2015). IRI/CPC ENSO Predictions Plume. Available at `http://iri.columbia.edu/our-expertise/climate/forecasts/enso/current/`, accessed on Sep. 23, 2015.

Magrath, J. (2014). Coffee rust fungus threatens employment collapse in central america.

Menke, W. (2012). *Geophysical data analysis: discrete inverse theory*. Academic press.

Modern Farmer (2014). Battling the coffee rust: Photos of farmers fighting an epidemic. Retrieved May 16, 2015, from `http://modernfarmer.com/2014/08/battlingcoffee-growers-struggleepidemic/`.

Monfreda, C., Ramankutty, N., and Foley, J. A. (2008). Farming the planet: 2. geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global biogeochemical cycles*, 22(1).

NCAR (2015). Ncep/ncar reanalysis monthly jeans and other derived variables. Retrieved 17 Apr. 2015 from `http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.html`.

Nelsen, R. B. (2013). *An introduction to copulas*, volume 139. Springer Science & Business Media.

NOAA Climate Prediction Center (CPC) (2015). Climate indices: Monthly atmospheric and ocean time series. Available at `http://www.esrl.noaa.gov/psd/data/climateindices/list/`.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al. (2010). The ncep climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1057.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of sciences*, 106(37):15594–15598.

Sweet Maria (2015). World coffee production timetable. Available at `http://sweetmarias.com/coffee.prod.timetable.php`.

Thurston, R. W., Morris, J., and Steiman, S. (2013). *Coffee: A Comprehensive Guide to the Bean, the Beverage, and the Industry*. Rowman & Littlefield Publishers.