All things plants:  
An ecosystem view of sustainable development

Joséphine Gantois

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2021
ABSTRACT

All things plants: An ecosystem view of sustainable development

Joséphine Gantois

Achieving societal well-being goals is inextricably linked to the preservation of many ecosystem functions. This dissertation adopts a plant lens, to contribute to our understanding of sustainable ecosystem functioning. Specifically, it sheds light on some plant physiology, phenology, and ecology processes, which matter for sustainable development: tree growth response to high temperatures, annual fluctuations in the timing of plant flowering, and ecological benefits of crop diversity that translate into economic returns. In addition, it illustrates how large-scale data proxies can be used to document large scale patterns that arise from individual plant processes. Chapter 1 documents a new methodology for estimating tree-level temperature response curves, using tree ring data and a degree-day framework. It uses those curves to document harmful impacts of high temperatures for tree growth across the US, and shows that there is limited acclimatization, but some adaptation to those high temperatures in a sample of climate sensitive and long-lived trees. Chapter 2 shows that satellite imagery and deep learning tools can be leveraged, to monitor interannual variations in the timing of plant flowering at large scales. It documents the predictive performance of two models: one adapted to monitoring crop flowering, the other adapted to monitoring shifts in the onset of spring flowering. Finally, chapter 3 highlights remaining gaps between empirical evidence of crop diversity benefits, and portrayal of those benefits in economic models of optimal crop diversity choice. Together, these chapters illustrate that bridging scales and disciplines is a difficult task, although it is necessary for understanding the sustainability of the human environmental footprint.
# Contents

List of Figures

List of Tables

Introduction

1 New tree-level temperature response curves document sensitivity, acclimatization, and adaptation to high temperatures
   1.1 Introduction ..................................................... 7
   1.2 Materials and methods ........................................ 10
      1.2.1 Tree growth data ......................................... 10
      1.2.2 Weather data ............................................. 10
      1.2.3 Ecoregion data ........................................... 12
      1.2.4 Temperature response curve ............................. 13
         1.2.4.1 Baseline model: response to year-round temperature 13
         1.2.4.2 Seasonal model: response to spring-summer temperature 15
         1.2.4.3 Controlling for drought ............................... 16
         1.2.4.4 Sample restrictions: temperature-sensitive trees, and isohydric versus anisohydric trees 16
         1.2.4.5 Lagged models ........................................... 17
      1.2.5 Applications ............................................... 17
         1.2.5.1 Acclimatization analysis ............................. 17
         1.2.5.2 Adaptation analysis .................................. 20
   1.3 Results .......................................................... 20
      1.3.1 Temperature response curve ............................. 20
         1.3.1.1 Baseline model: response to year-round temperature 20
         1.3.1.2 Seasonal model: response to spring-summer temperature 21
         1.3.1.3 Controlling for drought ............................... 22
1.3.1.4 Sample restrictions: temperature-sensitive trees, and isohydric versus
anisohydric trees ..................................................... 24
1.3.1.5 Lagged models .................................................. 26
1.3.2 Applications ......................................................... 27
  1.3.2.1 Acclimatization to high temperatures ......................... 27
  1.3.2.2 Adaptation to high temperatures .............................. 28
1.4 Discussion .......................................................... 31

A Appendix to Chapter 1 .................................................. 35
  A.1 Additional data ....................................................... 36
    A.1.1 Individual maps of Köppen-Geiger ecoregions ................. 36
    A.1.2 Map of dominant species ........................................ 37
    A.1.3 Map of isohydric and anisohydric species ...................... 38
  A.2 Additional methods ................................................ 39
    A.2.1 Tree growth data: formatting .................................... 39
    A.2.2 Tree growth data: list of corrections .......................... 40
    A.2.3 Temperature response curve: age control and standard errors ... 41
    A.2.4 Temperature response curve: spline model .................... 41
    A.2.5 Acclimatization analysis: choice of age cutoff characterizing tree youth ... 42
  A.3 Additional results .................................................. 43
    A.3.1 Temperature response curve: spline model .................... 43
    A.3.2 Temperature response curve: restricted range for optimal breakpoint search ... 44
    A.3.3 Temperature response curve, map, and species composition, by ecoregion .... 45
    A.3.4 Map of temperature-sensitive trees ............................ 51
    A.3.5 Temperature response curve: temperature-sensitive trees, controlling for drought ... 52
    A.3.6 Temperature response curve: splitting Eastern ecoregions by isohydric status .. 53
    A.3.7 Legacy effects of temperature: distributed lag model ........... 54
    A.3.8 Acclimatization: year-round temperature response .......... 56
    A.3.9 Acclimatization: AMJJA temperature response ................ 60
    A.3.10 Acclimatization: AMJJA temperature response of temperature-sensitive trees . 63
    A.3.11 Acclimatization: decadal model .............................. 64
    A.3.12 Adaptation: all humid ecoregions of the US, 23°C breakpoint ........ 65
# Monitoring shifts in flowering phenology using satellite imagery and deep learning

## 2 Introduction

**Introduction**

**Related Work**

**Core Approach**

- **Idea:** exploit convolutional neural networks and transfer learning
- **Model architecture**
- **Model evaluation**
- **Input data: satellite imagery**

## 2.1 Introduction

**Introduction**

## 2.2 Related Work

**Related Work**

## 2.3 Core Approach

**Core Approach**

- **Idea:** exploit convolutional neural networks and transfer learning
- **Model architecture**
- **Model evaluation**
- **Input data: satellite imagery**

## 2.3.1 Idea: exploit convolutional neural networks and transfer learning

**Idea: exploit convolutional neural networks and transfer learning**

## 2.3.2 Model architecture

**Model architecture**

## 2.3.3 Model evaluation

**Model evaluation**

## 2.3.4 Input data: satellite imagery

**Input data: satellite imagery**

## 2.4 Use Case 1: Predicting Crop Flowering Progress

**Use Case 1: Predicting Crop Flowering Progress**

- **Model**
- **Data**
  - **Ground level crop phenology**
  - **Weather data for transfer learning**
  - **Crop study region and summary statistics**
- **Results**

## 2.5 Use Case 2: Predicting the Onset of Spring Flowering

**Use Case 2: Predicting the Onset of Spring Flowering**

- **Model**
- **Data**
  - **Ground level flowering onset**
  - **Weather-based proxy for transfer learning**
  - **Study region and summary statistics**
- **Results**

## 2.6 Discussion

**Discussion**

## B Appendix to Chapter 2

**Appendix to Chapter 2**

- **Data**
  - **USDA NASS crop progress reports**
- **Additional figures**
  - **USDA NASS crop flowering progress**
  - **Observed and predicted date of 50% flowering (district-level)**
  - **Observed and predicted crop flowering progress (district-level)**
3 How do economic models of crop choice portray the benefits of diverse crop ecosystems?

3.1 Introduction

3.2 Background: various channels link crop diversity with increased crop yield, stability, and resilience

3.3 Potential sources of disconnect between ecological characteristics and economic modeling

3.3.1 Key characteristics of crop diversity benefits

3.3.1.1 Spatial feature: large scale spillover of crop diversity benefits

3.3.1.2 Spatial feature: large scale negative externality from substituting crop diversity benefits with human input substitutes

3.3.1.3 Temporal feature: lagged private benefits

3.3.1.4 Temporal feature: temporal (and spatial) externality via evolutionary forces

3.3.1.5 Low probability catastrophic events

3.3.2 Economic modeling of crop diversity choice

3.3.2.1 Theoretical model: optimal crop diversity level, assuming an explicit pathogen channel that has global spillovers

3.3.2.2 Theoretical model: optimal crop diversity level and tax on pesticides, assuming an implicit pest channel that has regional spillovers, and negative externalities from pesticide use

3.3.2.3 Theoretical model, with an empirical estimation: impact of neighboring crop diversity on productivity, through an implicit pest channel

3.3.2.4 Theoretical model, with an empirical estimation: impact of in situ crop diversity on productivity, through an explicit nutrient channel

3.4 Illustration: neighborhood effects and multiple equilibria

3.5 Conclusion

Bibliography
List of Figures

1.1 Map of tree ring sequences and frequency of measurements over time . . . . . . . . . . . . 11
1.2 Map of tree ring sequences, by Köppen-Geiger ecoregion . . . . . . . . . . . . . . . . . . 13
1.3 Acclimatization hypotheses . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
1.4 Piecewise linear relationship between annual log ring width and temperature . . . . . . 21
1.5 Piecewise linear relationship between annual log ring width and AMJJA temperature . . 22
1.6 Piecewise linear relationship between annual log ring width and temperature, controlling
for drought . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
1.7 Piecewise linear relationship between annual log ring width and AMJJA temperature, con-
trolling for drought . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
1.8 Piecewise linear relationship between annual log ring width and AMJJA temperature for
temperature-sensitive trees . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
1.9 Piecewise linear relationship between annual log ring width and temperature, controlling
for one lag in weather . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
1.10 Ecoregion-level acclimatization model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
1.11 US-wide acclimatization model, temperature-sensitive trees . . . . . . . . . . . . . . . . 29
1.12 Adaptation maps: frequency and impact of degree-days above 23°C in humid ecoregions
(selected grid cells) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
1.13 Adaptation scatterplot: impact of degree-days above 23°C as a function of their frequency,
in humid ecoregions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31
A.1 Individual maps of tree ring sequences, by Köppen-Geiger ecoregion . . . . . . . . . . . . 36
A.2 Map of tree ring sequences, by species . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
A.3 Map of tree ring sequences, by isohydric status . . . . . . . . . . . . . . . . . . . . . . . . . 38
A.4 Variation in tree growth as a function of tree age . . . . . . . . . . . . . . . . . . . . . . . 42
A.5 Flexible spline-based relationship between annual log ring width and temperature . . . . 43
A.6 Piecewise linear relationship between annual log ring width and temperature, restricted
search range for the breakpoint . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.13</td>
<td>Map of tree ring sequences and frequency of measurements over time, temperature-sensitive trees</td>
<td>51</td>
</tr>
<tr>
<td>A.14</td>
<td>Piecewise linear relationship between annual log ring width and AMJJA temperature for temperature-sensitive trees, controlling for drought</td>
<td>52</td>
</tr>
<tr>
<td>A.15</td>
<td>Piecewise linear relationship between annual log ring width and temperature, by isohydric status</td>
<td>53</td>
</tr>
<tr>
<td>A.16</td>
<td>Temperature coefficients from a distributed lag model</td>
<td>54</td>
</tr>
<tr>
<td>A.17</td>
<td>AMJJA temperature coefficients from a distributed lag model</td>
<td>55</td>
</tr>
<tr>
<td>A.18</td>
<td>AMJJA temperature coefficients from a distributed lag model for temperature-sensitive trees</td>
<td>56</td>
</tr>
<tr>
<td>A.19</td>
<td>Adaptation map: impact of degree-days above 23°C in humid ecoregions (all grid cells)</td>
<td>65</td>
</tr>
<tr>
<td>A.20</td>
<td>Adaptation maps: frequency and impact of degree-days above 26°C in region Cfa (selected grid cells)</td>
<td>66</td>
</tr>
<tr>
<td>A.21</td>
<td>Adaptation scatterplot: impact of degree-days above 26°C as a function of their frequency, in region Cfa</td>
<td>66</td>
</tr>
<tr>
<td>A.22</td>
<td>Adaptation maps: frequency and impact of degree-days above 17°C among Douglas Fir trees (selected grid cells)</td>
<td>67</td>
</tr>
<tr>
<td>A.23</td>
<td>Adaptation scatterplot: impact of degree-days above 17°C as a function of their frequency, among Douglas Fir trees</td>
<td>67</td>
</tr>
<tr>
<td>A.24</td>
<td>Adaptation maps: frequency and impact of degree-days above 14°C among Ponderosa Pine trees (selected grid cells)</td>
<td>68</td>
</tr>
<tr>
<td>A.25</td>
<td>Adaptation scatterplot: impact of degree-days above 14°C as a function of their frequency, among Ponderosa Pine trees</td>
<td>68</td>
</tr>
<tr>
<td>A.26</td>
<td>Adaptation maps: frequency and impact of degree-days above 23°C among White Oak trees (selected grid cells)</td>
<td>69</td>
</tr>
<tr>
<td>A.27</td>
<td>Adaptation scatterplot: impact of degree-days above 23°C as a function of their frequency, among White Oak trees</td>
<td>69</td>
</tr>
<tr>
<td>2.1</td>
<td>Snapshots of Planet data over the Carrizo Plain in California</td>
<td>77</td>
</tr>
<tr>
<td>2.2</td>
<td>Core model architecture</td>
<td>79</td>
</tr>
<tr>
<td>2.3</td>
<td>Crop model architecture</td>
<td>81</td>
</tr>
<tr>
<td>2.4</td>
<td>Crop model study region and sample data maps</td>
<td>84</td>
</tr>
<tr>
<td>2.5</td>
<td>Annual evolution of NDVI and temperature over a randomly selected crop field</td>
<td>85</td>
</tr>
</tbody>
</table>
2.6 Interannual variability in the date of 50% of flowering, by agricultural district and crop type  85
2.7 District-level timing of 50% crop flowering, observed and predicted .......................... 87
2.8 Pixel-level maps of 50% corn silking, “observed” and predicted ................................. 89
2.9 Pixel-level maps of 50% soybean blooming, “observed” and predicted ......................... 89
2.10 Spring onset model architecture ..................................................................................... 91
2.11 Spring onset model study region ..................................................................................... 93
2.12 Temperature-based first bloom index over the core study region: observed and predicted . 94
2.13 Variability in model performance across the core study region ..................................... 96

B.1 USDA NASS corn silking progress, by agricultural district and year ............................ 101
B.2 USDA NASS soybean flowering progress, by agricultural district and year ................... 102
B.3 District-level timing of 50% corn silking, observed and predicted (interannual variations by
   district) ................................................................................................................................. 103
B.4 District-level timing of 50% soybean flowering, observed and predicted (interannual varia-
   tions by district) .................................................................................................................... 104
B.5 District-level corn silking progress, observed and predicted ........................................... 105
B.6 District-level soybean flowering progress, observed and predicted .............................. 106
B.7 Interannual variability in the first bloom index ............................................................... 107
B.8 Variability in model performance across the Eastern US .............................................. 108
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)</td>
<td>57</td>
</tr>
<tr>
<td>A.2</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)</td>
<td>57</td>
</tr>
<tr>
<td>A.3</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)</td>
<td>58</td>
</tr>
<tr>
<td>A.4</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)</td>
<td>58</td>
</tr>
<tr>
<td>A.5</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)</td>
<td>59</td>
</tr>
<tr>
<td>A.6</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)</td>
<td>59</td>
</tr>
<tr>
<td>A.7</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)</td>
<td>60</td>
</tr>
<tr>
<td>A.8</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)</td>
<td>60</td>
</tr>
<tr>
<td>A.9</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)</td>
<td>61</td>
</tr>
<tr>
<td>A.10</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)</td>
<td>61</td>
</tr>
<tr>
<td>A.11</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)</td>
<td>62</td>
</tr>
<tr>
<td>A.12</td>
<td>Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)</td>
<td>62</td>
</tr>
<tr>
<td>A.13</td>
<td>Influence of early exposure to degree-days among T-sensitive trees (location FE in the interaction)</td>
<td>63</td>
</tr>
</tbody>
</table>
Acknowledgements

This dissertation would not exist without my wonderful group of advisors, who supported and guided me throughout. I would like to thank in particular Jeffrey Shrader, for his steadfast enthusiasm and meticulous feedback, across the disparate pieces of my research agenda. I am grateful to have been your teaching assistant on the challenging but immensely rich Human Ecology course. Preparing for and teaching this class has triggered many clarifying moments. I would like to thank Wolfram Schlenker, for planting the seeds of what has become my first successful research project. I am grateful for your unfailing encouragement, for giving me confidence, when I doubted my ideas and results, that they were worth pursuing, and for your rigorous feedback.

On the western side of Amsterdam avenue, over at E3B, Duncan Menge and Ruth Defries provided me with a second home, by welcoming me in their warm and intellectually rich lab groups. The weekly check-ins and constant humor I found in Duncan’s lab have been a life saver at times. I look up to your qualities as a mentor, a teacher, and the humble conductor of an inclusive and incredibly supportive research lab. I would also like to thank Ruth Defries, for giving me the courage to knock at different E3B lab group doors; this has profoundly impacted my research path, for the better I believe. I have been and continue to be inspired by your lab group’s research and the high standards you hold all of us to, always making sure that we engage in impactful research, that we don’t lose track of the ground, and that we carefully involve local expertise. You are the keeper of my academic moral compass.

Many more faculty guided me along my research journey, including Doug Almond, Brendan Buckley, Edward Cook, Donald Davis, Kevin Griffin and his lab, Supreet Kaur, Amit Khandelwal, Claire Kremen and her lab, Upmanu Lall, Suresh Naidu, and Maria Uriarte. Some also helped me grow as a teacher, particularly Ruth Defries, John Mutter, Jeff Shrader, Miguel Urquiola, and three patient cohorts of MPA-ESP students, who helped me craft a better version of my intensive summer math review. The Center for Teaching and Learning at Columbia has been the backbone of my teaching journey. I leave Columbia with invaluable teaching and mentoring skills, and great examples of how to bring together a truly interdisciplinary crowd. And of course, Tomara Aldrich has been immensely helpful for taking all administrative issues off of my shoulders and more, and for being such a sensible advocate and source of support in difficult times.
The big Sustainable Development family, past and present, is the reason why I joined this PhD program in the first place. From the now alumni, who opened their colloquium and seminar doors while I was just an applicant, to the most recent cohorts, who are bringing great energy and enthusiasm into the program, I have received much support, intellectual and emotional, throughout the past six years. A special thanks to Sandra and Raimundo, the best cohort mates I could have asked for, for their precious cheerfulness and relaxed attitude. I owe much of my sanity to Ana, Helen, Ruiwen, and Chas, and Kimberly and Aaron before them, for pretending alongside me that combining kids and a PhD is a perfectly reasonable mix. And to Anouch, Maya, Thomas, and many friends beyond Columbia, for taking such good care of mine. I received wonderful research and non-research help (think volley ball, pasta making, burlesque shows, and card games) from many more.

Thank you Mathias, for teaming up and sticking with me in our crazy adventure of getting PhDs with minimal child care, for being an endless source of ideas, and for your invaluable proofreading. Thank you Zélie and Eléa for always being there, quite literally, throughout my PhD. You have attended many more meetings and seminars than you had asked for, and you always did so with such grace. Your candid questions and our slow-paced nature walks, have shaped my research pursuits more than you can imagine. Thank you to our dear families and friends in France, who took care of us from near and far, and always supported our goals.

Last, but not least, this dissertation work would not have been possible without the many excellent resources made publicly available. I am particularly indebted to R, Python, the International Tree Ring Data Bank, Google Earth Engine cloud computing power and free data archives that include quality calibrated products, Wolfram’s weather data extension from PRISM, and many ecology data repositories like NEON and TRY.
To Eléa and Zélie, with much love and hope for the future.
Introduction

- the songs of plants
  
  *a new ring crackles*
  
  *as colors pop*
  
  *ancient voices rise*

Understanding the environmental footprint of humans is a central concern for anyone wondering about the sustainability of our presence and way of living on Earth. This footprint is increasingly apparent when it comes to biodiversity and ecosystems. The most recent Global Assessment on Biodiversity and Ecosystem Services (Díaz et al., 2020), orchestrated by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), illustrates just that. Through careful examination of the state of Nature and its contribution to people, this report highlights the ongoing deterioration of biodiversity and ecosystem functions and services, many of which are key for maintaining and improving societal well-being goals.

One way to examine our environmental footprint is to assess to what extent changes in biodiversity and ecosystem function can be attributed to humans. A question arises: how do we understand the place of humans within ecosystems? If we go the anthropogenic attribution route, what is the null model, against which we are contrasting the current state of the world? This is tricky to answer. For one, humans are an integral part of ecosystems and can be sustainably so (Crabtree et al., 2019). In addition, nature is inherently dynamic. Regular perturbations like fires or droughts, or longer-term evolutionary adaptation, are essential to ecosystem functioning (DeFries and Nagendra, 2017). So ecosystem change is the result of intertwined non-anthropogenic and anthropogenic forces. Monitoring human impacts on ecosystems thus requires either to understand non-anthropogenic changes well enough to separate them from the non-sustainable human component (for example recent species extinction rates are compared to an estimated “background” pre-industrial extinction rate in Ceballos et al., 2015), or to forego the human-nature distinction entirely and focus on monitoring the functional sustainability of ecosystems. This ecosystem lens is the focus that I choose for my dissertation.

---

1 Another approach is to contrast the biodiversity and ecosystem function implications of different policy choices.
specifically, I seek to shed light on plant physiology, phenology, and ecology processes, which matter for sustainable development: tree growth response to high temperatures, annual fluctuations in the timing of plant flowering, and ecological benefits of crop diversity.

From an economic standpoint, my work falls under the “strong” version of sustainability. Weak sustainability focuses on maintaining a constant per capita level of aggregate capital stock, with the central assumption that all types of capital (physical, social, human, and natural) are substitutable, thanks to technological innovation (Hackett et al., 2006). Strong sustainability on the other hand emphasizes the critical role played by ecosystems and gives particular attention to functional integrity. My work contributes to the latter conceptualization, by making plant processes the central outcome.

A common driver of my work is to draw large scale patterns from individual plant processes. Plant measurements are costly to acquire at large spatial and temporal scales. I try to scale observations about plant response by exploiting large-scale data proxies: a public repository of tree ring data in chapter 1, and remotely sensed reflectance data in chapter 2. Relatedly, theory is lacking to explain how macroscopic patterns emerge from microscopic patterns in complex biological systems (Levin, 2005). I don’t address this gap, but I try to be mindful about how the large scale patterns I draw connect with our understanding at smaller scales, whether it is sub-tree physiological processes, individual plant phenology, or local plant-pest dynamics.

My work is also motivated by my perception that there is much that economic models and policy decisions overlook, in particular when it comes to ecosystem processes, for want of firm quantification or understanding (actual or perceived). At the same time, a lot of risks are associated with how these processes are reacting to anthropogenic pressure, such as risks of non linear high heat impacts, ecosystem collapse, or widespread pest-driven crop failures (DeFries et al., 2019). My work attempts to improve or review our understanding of some plant processes, for fairer inclusion in economic and policy discussions.

Chapter 1: New tree-level temperature response curves document acclimatization and adaptation to high temperatures. In this chapter, I document how temperature impacts tree growth non-linearly, with particularly harmful impacts of high temperatures, not estimated before. I construct precise temperature curves across ecoregions of the US, by combining tree ring data and a

---

This is interesting if the set of policy choices is well defined, though often constrained by data availability. It does not require to define a null model.

2This is likely one key factor underlying the lack of interaction between economics and ecology.
cumulative degree day model. These curves reflect both direct temperature impacts, via enzymatic activity rates for example, and indirect temperature impacts, via coupled stressors like drought. I find that the optimal temperature level for tree growth is around 20-25°C in humid ecoregions, while the lower temperature breakpoints of 10-15°C in dry ecoregions partially reflect strong drought influence. I isolate the direct temperature impacts by constructing a temperature response for the restricted sample of temperature-sensitive trees: their spring-and-summer temperature optimum is 32°C, which is in line with experimental evidence. In terms of high temperature impact, I find that six additional degree-days above the regional optimum leads to an average decrease in annual tree ring width of about 1%, with variations across ecoregions and seasons. I exploit those temperature curves to test evidence of medium-term acclimatization and long-term adaptation to supraoptimal temperatures. I find limited evidence of acclimatization: weather experienced during youth does not substantially influence a tree’s ability to withstand high temperatures later in life, relative to its neighbors. If anything, among temperature-sensitive trees, high temperature experience decreases a tree’s ability to withstand high temperatures. I find some evidence of adaptation: across humid ecoregions of the US, high temperatures are 43% less harmful to tree growth, in locations where the average incidence of high temperatures is one standard deviation above the average. This work bridges leaf-level estimates and ecosystem-level estimates of temperature optima and long-run adaptation. It is the first to document tree acclimatization to heat stress in a real world setting, by exploiting variation in individual tree histories.

Chapter 2: Monitoring shifts in flowering phenology using satellite imagery and deep learning. In this chapter, coauthored with Mathias Lécuyer, we ask whether satellite imagery can be leveraged to construct large scale measures of the timing of plant flowering. Our intuition is that some flowering events manifest via sharp transitions in spectral reflectance, which are nonetheless subtle, varied, and difficult to specify a priori. We turn to deep learning to extract temporal patterns in spectral reflectance that capture those transitions. We overcome the issue of scarcity of ground truth data by first training our model to predict a temperature proxy for flowering, and then fine-tuning our model parameters on more restricted ground-level observations of plant flowering. We illustrate this approach with two use cases. First, over cropland in Illinois and Iowa, we construct a proxy measure of the evolution of crop flowering—silking for corn, flowering for soybean—across the growing season. This measure improves on a null average model and estimates from the literature, and captures some, though not all, of the interannual variability in the timing of crop flowering. Second, over the Eastern
US, we construct a proxy measure of the onset of spring flowering, which reflects the phenology of early-flowering, temperature-sensitive species. This measure improves over the current temperature-based first bloom index from the National Phenology Network, and extrapolates much better over space. Again, it captures some, though not all, of the interannual variability in the onset of spring flowering. Our proposed approach, despite using coarse satellite inputs and restricted ground truth data on flowering, holds great promise as more of those data sources become available.

Chapter 3: How do economic models of crop choice portray the benefits of diverse crop ecosystems? In this chapter, I am more explicit about the societal implications of ecosystem functioning. Specifically, I consider crop species diversity in agricultural systems, and examine whether economic models of crop diversity choice fairly reflect the documented impacts of crop diversity on yield. I start by reviewing the main mechanisms through which crop diversity can increase yield in agricultural systems, such as the provision of natural pest control. Building on this review, I identify several specificities of crop diversity benefits (related to spatial externalities, temporal externalities, and the presence of low probability catastrophic events), which matter for the derivation of socially optimal levels of crop diversity. I highlight how existing economic models of crop choice reflect some of these specificities, and how there remains gaps between our ecological understanding and economic modeling of crop diversity benefits. This review sets the stage for improved economic modeling of crop diversity choice, which has implications for discussions of agricultural and trade policies.
Chapter 1

New tree-level temperature response curves document sensitivity, acclimatization, and adaptation to high temperatures
Abstract

Temperature is a key climate indicator, whose distribution is expected to shift right in a warming world. However, the high temperature tolerance of trees is less widely understood than their drought tolerance, especially when it comes to sub-lethal impacts of temperature on tree growth. In this chapter, we use a large data set of tree ring widths to estimate the relationship between temperature and tree radial growth across different ecoregions of the US. In particular, we demonstrate the performance of a flexible degree-day framework for documenting temperature impacts on tree radial growth, which overcomes the limitations of using coarser annual temperature aggregates. We find that tree radial growth responds non-linearly to temperature across all ecoregions of the US: temperature increases are beneficial or mostly neutral for tree growth up to around 20-25°C in humid regions and 10-15°C in dry regions, beyond which temperature increases are harmful to tree growth. Six additional degree-days above the local optimal temperature lead to an average decrease in tree ring width of around 1%, with variation across ecoregions and seasons. We show that, in dry regions particularly, the low temperature threshold partially reflects the influence of temperature-mediated drought, winter temperature response, and lagged weather impacts. The influence of other factors like soil moisture, nutrients or pests remains to be explored. We isolate direct temperature impacts by constructing a spring-and-summer temperature response for the restricted sample of temperature-sensitive trees: their critical threshold is higher, at 32°C, which is in line with experimental evidence. We then exploit local variation in the timing of tree birth, which determines the weather sequence experienced over a tree’s life, to show limited evidence of acclimatization. Among temperature-sensitive trees, greater experience of high temperatures early in life is associated with a slightly more severe impact of high temperatures later in life. This pattern is consistent with high temperature events damaging a tree’s structure hence its ability to limit growth fluctuations under high temperature events. Additionally, we find some evidence of long-term adaptation of trees to regional climates: across humid ecoregions of the US, high temperatures are 43% less harmful to tree growth, in locations where the average incidence of high temperatures is one standard deviation above the average. Overall, these results highlight the strength of a new methodology which, applied to more representative tree ring data, could help predict tree carbon uptake potential and forest composition under global change.
1.1 Introduction

Temperature exerts a strong influence on tree growth, enhancing it at some levels, while harming it at others. Resolving the temperature response of tree growth is important in a context where ongoing warming is expected to shift the full temperature distribution upwards (Pachauri et al., 2014), with consequences for forest carbon uptake that will drive much of the carbon cycle response to climate extremes (Frank et al., 2015).

There is a dominant focus in the literature on the impact of drought on tree growth, while the impact of high temperatures is less widely understood (Breshears et al., 2021; Allen et al., 2015). We argue that focusing on temperature is valuable: temperature is a key climate indicator; it underlies many impacts on tree growth, both direct (impacts on photosynthesis, respiration, and cambial phenology among others) and indirect (through its coupling with soil moisture, vapor pressure deficit, or solar radiation); and the focus on a single factor allows flexible and scalable modeling, compared to a data-intensive process-based approach. Specifically, we propose a degree-day model, combined with tree ring data, to resolve the real-world temperature response of annual tree radial growth at large scales. We test this approach on open-source tree ring data to highlight its potential and suggest avenues for future work.

Our tree-level temperature response curves exploit the flexibility of degree-days. There is an extensive literature documenting the relationship between individual tree growth and temperature, in part for the purpose of reconstructing historical weather from long tree ring records (Fritts, 2012). Large scale studies usually rely on annual or seasonal temperature aggregates, such as average temperature, or maximum temperature over the growing season, to match the temporal resolution of tree ring data. This type of aggregation masks within-year fluctuations, including the intensity and severity of extreme events, and prevents the estimation of a precise temperature response curve across the full range of possible temperatures. Our work aggregates temperature at the annual level using a cumulative degree-day formulation: each degree experienced during the year contributes to that year’s growth, and each degree can have a different impact on growth depending on the temperature bracket it belongs to. This framework is common in agronomic models. For example, it has been used to estimate the temperature response curve of different crops, and highlight their sensitivity to high temperatures especially (Schlenker and Roberts, 2009). It is also present in tree-specific studies: growing degree-days have been used to predict the timing of cambial reactivation at the beginning of the growing season (Begum et al., 2018), and growing degree-days have been reconstructed from historical tree ring data.
at Northern latitudes (Jacoby et al., 1985).

Our degree-day model highlights the sub-lethal effects of temperature: there exists an “optimal” temperature for tree radial growth, much lower than the critical thermal limit of trees, beyond which growth decreases with temperature. This work bridges the leaf-level and ecosystem-level literatures, which also suggest a non-linear response of tree growth to temperature. At the leaf level, a large experimental literature clearly points to the existence of an optimal temperature for net photosynthesis, based on direct temperature effects on photosynthesis and respiration. This optimum often falls between 25°C and 35°C, though with large variations across climatic regions (Kattge and Knorr, 2007; Lloyd and Farquhar, 2008; Medlyn et al., 2002; Waring and Running, 2010). At the the ecosystem level, a recent paper documents the existence of a similar temperature optimum for gross primary productivity, which primarily ranges between 15°C and 30°C across the US (Huang et al., 2019). Their estimate reflects real-world conditions, and thus incorporates both the direct effects of air temperature, and indirect effects through atmospheric vapor pressure deficit (VPD), soil moisture, and solar radiation. The relative influence of each channel is not parsed out, though the authors show that controlling for VPD does not fully crowd out the influence of air temperature on the temperature optimum.

Overall, our paper is the first to construct real-world temperature response curves for tree radial growth at a large scale. It bridges the experimental leaf-level and tree-level views with the real-world ecosystem perspective, and discusses differences in response across ecoregions. More importantly, our paper documents both the optimal temperature for growth, and the harmful impact of supraoptimal temperatures, which is key for assessing the impact of shifting temperature distributions on tree radial growth. While our approach does not disentangle the role of underlying physiological dynamics, we explore how our temperature response curves connect to the physiological literature. In particular, we test how much atmospheric drought drives temperature effects, and further isolate direct temperature effects by exploiting the restricted sample of trees whose limiting growth factor is temperature and not drought.

We exploit our newly drawn temperature response curves to document whether trees acclimatize to high temperatures. Specifically, we document how a tree’s temperature sensitivity is shaped by its previous exposure to high temperatures and variable weather. Such phenotypic plasticity is likely to play a crucial role in shaping tree persistence and forest composition under rapid climate change, given the slower pace of evolutionary and migratory responses (Gratani, 2014). There is evidence of short-run acclimation in experimental settings (Nievola et al., 2017). Day-long to year-long warming induces a
rightward shift of photosynthesis response to temperature (Kattge and Knorr, 2007; Gunderson et al., 2010; Wang et al., 1996). Respiration fully acclimates, such that there is an upward and rightward shift of the net photosynthesis response to temperature, under warming conditions (Smith and Dukes, 2013). The recent ecosystem level study did not find evidence of acclimatization to the 1°C warming trend of the past three decades (Huang et al., 2019). Our paper provides a test of medium-run acclimatization of tree growth to high temperatures. It relies on differences in early life weather exposure, among trees living in the same location. Specifically, we test the following hypotheses:

1. Trees acclimatize to high temperatures over their lifetime: their adult response is influenced by weather exposure experienced during their youth.
2. Tree acclimatization manifests via a milder impact of high temperatures, for trees more frequently exposed to high temperatures during their youth.
3. Tree acclimatization manifests via a more severe impact of high temperatures, for trees more frequently exposed to high temperatures during their youth.

The current distribution of trees is the result of long-term evolutionary adaptation to the local environment. We make use of the large geographical span of our tree ring dataset to document long-run adaptation of tree growth to high temperatures. There is some within-species evidence that trees adapt to their long-term local climate, based on their growth response to mean annual temperature (Canham et al., 2018). Evidence on the evolution of high heat tolerance is more scarce. O’sullivan et al. (2017), use upper canopy leaf level measurements across several biomes, to show that high temperature tolerance increases from the pole to the equator, though not as much as the difference in high temperature extremes would suggest. Similarly, Huang et al. (2019), find that ecosystem-level optimal temperature increases with maximum temperature over the growing season, although the slope of this relationship is lower than 1. We test the following hypothesis:

4. Regional temperature response curves reflect long-term evolutionary adaptation: temperature optima are higher and the impact of supra-optimal temperature lower in climates characterized by more frequent high temperatures.
1.2 Materials and methods

1.2.1 Tree growth data

We exploit publicly available tree ring records to get data on tree radial growth. We extract all records voluntarily contributed to the International Tree-Ring Data Bank (ITRDB), and filter the data to restrict our study region to the contiguous US, and study period to 1901-2016. We use the most consistently reported measurement, total annual ring width, as our metric of tree growth. The resulting dataset has a panel structure, with yearly ring width measurements for each tree core. It contains 4,330,241 ring width measurements, from 51,900 tree cores, across 1284 locations. Measurements usually extend from the outermost ring under the bark to, or very near, the pith of the tree. There can be gaps due to true missing rings or measurement issues. In appendix A.2.1, we describe in detail how we processed the data and addressed formatting issues to make this public repository exploitable and reliable. Figure 1.1 shows the location of tree ring sequences and the sample size at each location. The full dataset, shown on the left, is used to construct temperature response curves. A smaller dataset, restricted to trees born on or after 1900, and shown on the right, is used to test the acclimatization hypotheses. These hypotheses require us to characterize individual trees' weather history from birth, and weather data are available from 1900 onwards, hence the data restriction.

The ITRDB does not represent a regionally representative sample of trees. It primarily contains data on trees used for reconstruction of climatic variables in the past, which are precisely selected for their responsiveness to climate. In fact, the climate sensitivity of these trees has been estimated to be about 50% higher than regional forest climate sensitivity in the South-West, based on a comparison with tree ring data from the Forest Inventory and Analysis program (Klesse et al., 2018). This does not affect the relevance of constructing temperature response curves for those climate-sensitive trees, or of testing the extent of their temperature acclimatization and adaptation. However, this prevents extrapolation to other tree species or forest-level dynamics, and might impact comparisons across regions if sampling strategies vary by location.

1.2.2 Weather data

We use temperature and precipitation data from the “Parameter-Elevation Regressions on Independent Slopes Model” (PRISM). Total precipitation is available at a monthly resolution across a 4 km x 4 km grid covering the contiguous US. For minimum and maximum temperature, monthly PRISM data are downscaled at the daily level across the entire PRISM grid, using a robust methodology
first implemented by Schlenker and Roberts (2006), and updated in 2020. It is essentially a three-step process: first, monthly data in each PRISM grid cell are regressed on monthly averages from the closest ten instrumental stations that have continuous coverage (the R$^2$ of this model is 0.999); second, missing daily values at those stations are interpolated from neighboring stations with non-missing values; third, the relationship estimated in the first step is used to infer daily values at each grid cell from daily values at the station-level. From these daily maximum and minimum temperature data, the within-day temperature distribution is extrapolated by fitting a sinusoidal curve. The end-product is a half-hourly temperature distribution, over 1900-2016, at a 4km resolution across the contiguous US. Given the extent of within-day temperature fluctuations, this last step allows to capture extreme high degree days much more precisely than when using daily average temperature. A key characteristic of the PRISM interpolation model is that it takes elevation into account, in addition to reproducing gradients caused by temperature inversions, rain shadows or coastal effects. Tree ring data cover large elevation ranges, especially in the South-West of the US, which makes PRISM data particularly adapted to our setting.

We want to estimate the total impact of temperature on tree growth, which combines direct effects like temperature-induced acceleration of enzymatic processes, and indirect effects via coupled factors like vapor pressure deficit and soil moisture. We thus do not systematically control for those coupled
factors in our main analyses. However, we use the Standardised Precipitation-Evapotranspiration Index (SPEI) in additional analyses, to test how much the temperature response is driven by temperature-mediated drought conditions. This is particularly relevant given the sampling bias in our data towards drought-sensitivity (Williams et al., 2013). The SPEI index captures deviations from the local average water balance (precipitation net of potential evapotranspiration). It is computed at different timescales. We use the 1-month and 3-month SPEI to capture short-term drought conditions, and the 12-month SPEI to capture medium-term drought conditions, which reflect both sustained atmospheric drought conditions, and slower evolving soil moisture conditions. We choose this index over other drought indices for its independence of tree ring data, ability to capture ecological impacts, and flexibility of time scales for capturing a priori complex and potentially heterogenous impacts across tree species and regions (Vicente-Serrano et al., 2013; Bhuyan et al., 2017).

1.2.3 Ecoregion data

We split the continental US into ecoregions. We use the Köppen-Geiger climate classification, which is driven by differences in vegetation types across locations. Each ecoregion is identified by a three-letter code. The first letter denotes the main climatic group: B for dry, C for temperate, and D for continental. Each climatic group is further divided into subgroups based on seasonal precipitation and temperature patterns: the second letter denotes the precipitation pattern (W for desert, S for steppe, f for fully humid, s for summer dry), and the third letter denotes the temperature pattern (k for cold arid, a for hot summer, b for warm summer, c for cool summer).

Figure 1.2 shows the location of tree ring sequences and the sample size at each location, by Köppen-Geiger ecoregion. Individual maps are displayed in appendix A.1.1, to further highlight geographical clustering in each ecoregion. We focus on ecoregions that span at least 35 different PRISM grid cells. This safeguards against low levels of variation in the temperature predictors, since temperature only varies by PRISM grid cell, and excludes less than 10% of the data. Two ecoregions are characterized by a dry climate: BSk in the South-West, and Csb along the Western coast. All other ecoregions are characterized as “fully humid” (“f”), with varying levels of summer temperatures (“a”, “b”, or “c”, in decreasing order of heat).

In the ITRDB data, tree species are very correlated with ecoregions: species selection is influenced by the local climate, and often, a single tree species is sampled at a given location. Figure A.2 in appendix A.1.2 shows the location of tree ring sequences and the sample size at each location, for the top ten tree species. Most species distributions align well with specific ecoregions, and there is
1.2.4 Temperature response curve

1.2.4.1 Baseline model: response to year-round temperature

We model the relationship between ring width and temperature using a piecewise linear function. We have a particular interest in the breakpoint where the slope of the relationship changes, which we expect to correspond to the optimal temperature level, and in the slope value beyond this breakpoint, which captures the impact of temperature increases at high temperature levels. We log transform tree
ring width to reflect its lognormal distribution. Temperature is captured by two explanatory variables in the model: total degree-days below the optimal temperature level and total degree-days above the optimal temperature level. This choice reflects an important assumption: temperature effects on tree growth are cumulative over time and annual tree growth is proportional to total exposure. Splitting degree-days into two variables allows temperature impacts to differ below and above the optimal temperature breakpoint. This optimal breakpoint is not known a priori, it is selected among all possible degree bounds in the \([0-40^\circ C]\) range, using a 10-fold cross-validation procedure that uses random sets of years to measure each model’s predictive performance. The bound that gives the lowest root mean squared error on average across all folds is selected as the optimal temperature breakpoint. We thus model log ring width, \(y_{ct}\), from tree core \(c\), formed in year \(t\), and located in PRISM grid cell \(s\), as:

\[
y_{ct} = \alpha_c + \beta_{1}\text{low\_degree\_days}_{st} + \beta_{2}\text{high\_degree\_days}_{st} + \delta z_{ct} + \epsilon_{ct}
\]

\[
\text{low\_degree\_days}_{st} = \left[ \sum_{\text{days } d \text{ in year } t} \frac{1}{48} \left( \sum_{\text{half-hour } h \text{ in day } d} \max(0, \min(T_{sh}, \bar{t})) \right) \right]
\]

\[
\text{high\_degree\_days}_{st} = \left[ \sum_{\text{days } d \text{ in year } t} \frac{1}{48} \left( \sum_{\text{half-hour } h \text{ in day } d} \max(0, T_{sh} - \bar{t}) \right) \right]
\]

Tree core fixed effects\(^1\) \(\alpha_c\) account for time-invariant characteristics of a tree that might be correlated with temperature: soil quality, baseline climate, species-specific average growth rate, etc. Including those fixed effects implies that the estimation exploits year-to-year variations in temperature, and does not rely on baseline temperature differences across locations, which helps isolate the impact of temperature on tree growth. The term \(z_{ct}\) includes two additional controls. The first is a quadratic in total annual precipitation, which allows for a non-linear impact of very low and very high precipitation levels. This controls for unusually wet or dry years, which might occur independently from high temperature events, and have a documented impact on tree ring width across the US (George, 2014). The second is a linear control in tree age, which allows for a linearly decreasing log growth rate over a tree’s lifetime, similar to a traditional form of age standardization (Schofield et al., 2016). Standard errors are clustered at a 2° gridcell level, to allow for spatial correlation in the error term\(^2\).

\(^1\)This terminology is used in the econometric sense.

\(^2\)See appendix A.2.3 for more details on the age control and clustering of standard errors.
The coefficient $\beta_1$ represents the impact of an additional degree-day below $\bar{t}$, and the coefficient $\beta_2$ represents the impact of an additional degree-day above $\bar{t}$.

To represent this model graphically, we vary the temperature for 1 day during any given year, between 0°C and 40°C. According to the model, the resulting change in log tree ring width is $\hat{\beta}_1 \text{low\_degree\_days}(T_{sd}) + $ $\hat{\beta}_2 \text{high\_degree\_days}(T_{sd})$, where $d$ is that fictitious day whose temperature we are changing. We plot that change as a function of $T_{sd}$. The temperature curve obtained represents the impact of a given day’s temperature, relative to a temperature of 0°C for that day.

In appendix A.2.4, we describe a more flexible spline-based model in temperature, similar to what has been used to study the temperature response of crops (Schlenker and Roberts, 2009). We use this flexible model to test whether the growth–temperature relationship emerges as piecewise linear, even without imposing a restrictive functional form. We do not use this model widely, because cross-curve comparisons are difficult, and because high order spline curves can behave wildly at the extremes, whereas we have a particular interest in the impact of extreme high temperatures (Albouy et al., 2016).

### 1.2.4.2 Seasonal model: response to spring-summer temperature

The baseline model averages the impact of a given temperature exposure across all seasons within a year. This could mask heterogeneity in seasonal temperature responses, in the temperature ranges that overlap across seasons. For instance, the temperature response curve in the mid-temperature range could be driven both by the impact of a warmer winter and by the impact of a warmer spring. At high elevations in dry climates, the former would translate into lower growth, due to thinner snowpack, hence lesser recharge of spring moisture, while the latter would translate into higher growth. Mixing seasonal temperature responses could also impact the temperature response outside of the range where seasonal temperatures overlap, given the low degrees of freedom of our piecewise linear functional form.

We are particularly interested in the impact of high temperatures, so we estimate a spring-summer temperature response. We use the structure of the baseline model, except that all weather variables—degree-days, precipitation, and drought when included—are summed over the spring-summer season rather than the full calendar year. We define the spring-summer season as the months of April through August (AMJJA).
1.2.4.3 Controlling for drought

Our baseline model captures the impact of temperature changes, inclusive of associated changes in coupled factors. One such factor, drought, is likely to play a dominant role in shaping the temperature-response curves, given strong drought sensitivity in our tree ring sample, especially in dry ecoregions BSk and Csb. In order to understand how much our temperature response curves are driven by temperature-mediated drought conditions, we estimate a "drought-free" temperature response by controlling for drought in our model. We include monthly values of the 1-month, 3-month, and 12-month SPEI, to capture both short-term and medium-term drought conditions, and control for drought flexibly. We include all months from January to December for the 1-month SPEI; the months of March, June, September, and December for the 3-month SPEI; and the month of December for the 12-month SPEI. Since the x-month SPEI takes into account drought conditions in the previous x months, this avoids excessive correlation across monthly SPEI values, and matches the calendar year that is used to construct the temperature and precipitation variables.

1.2.4.4 Sample restrictions: temperature-sensitive trees, and isohydric versus anisohydric trees

Although drought is the primary growth factor for many trees in the ITRDB dataset, some trees primarily respond to temperature. We call these the “temperature-sensitive” trees. We identify them using the raw correlation between log ring width and cumulative AMJJA temperature, and the criteria that this correlation be strictly positive (George, 2014). We expect the temperature response of these temperature-sensitive trees to reflect mostly direct temperature impacts. We estimate their spring-summer temperature response curve by fitting the AMJJA model to this restricted sample of trees.

We also estimate the temperature-response curve of trees by isohydric status. Isohydric trees tend to close their stomata when soil moisture decreases or VPD increases to avoid transpiring, while anisohydric trees adopt a riskier behavior and keep their stomata open longer under dry conditions. We expect the isohydric temperature response curve to show a milder response to high temperatures, compared to the anisohydric curve, due to the greater stomatal control of isohydric tree species. We assign an isohydric status to all oak and juniper species, and an anisohydric status to all pine, spruce, and tulip species. Despite strong geographic clustering of tree species in the ITRDB dataset, there is

\[ \text{This does not imply that temperature does not affect the other trees, just that it is not the primary factor modulating their growth.} \]
some spatial overlap between isohydric and anisohydric tree species in three ecoregions: Csb in the West, Dfb in the North-East and South-West, and Cfa in the South-East (see map in appendix A.1.3). We fit the baseline temperature model to isohydric and anisohydric trees separately, within each of those ecoregions.

1.2.4.5 Lagged models

Weather can be correlated from one year to the next, and last year’s weather can impact current tree radial growth via impacts on carbon uptake, root and xylem structures, or the timing of dormancy. Our baseline temperature response curves could thus partially reflect lagged temperature impacts. We estimate a lagged model, where we control for a single lag of all weather variables: splines in temperature, quadratic in precipitation, and drought variables when drought is included. The model is as follows:

\[
y_{ct} = \alpha_c + \beta_1 \text{low\_degree\_days}_{st} + \beta_2 \text{high\_degree\_days}_{st} + \delta z_{ct} + \gamma w_{c,t-1} + \epsilon_{ct}
\]

where the additional term \( w_{c,t-1} \) includes the weather lags described above, and the optimal breakpoint \( \bar{t} \) is again selected in the [0-40°C] range, using a 10-fold cross-validation procedure.

Trees store resources, in particular carbon, such that the carbon that they use to grow a new ring can come from carbon acquired in previous years. This means that past weather can have a strong impact on current growth. In addition, trees are long lived, and can react to a harmful weather event over several years, possibly compensating for growth failures by additional growth in later years (DeSoto et al., 2020). We document lagged impacts of supraoptimal temperatures by estimating a distributed lag model with 10 lags:

\[
y_{ct} = \alpha_c + \sum_{i=0}^{10} \beta_{1,i} \text{low\_degree\_days}_{s,t-i} + \sum_{i=0}^{10} \beta_{2,i} \text{high\_degree\_days}_{s,t-i} + \delta z_{ct} + \epsilon_{ct}
\]

where the optimal breakpoint \( \bar{t} \) is fixed at the value given by the baseline (no lag) model.

1.2.5 Applications

1.2.5.1 Acclimatization analysis

We use our temperature model, to test whether the trees’ long run response to temperature varies, as a function of the weather sequence experienced during their youth. We formulate two competing hypotheses. The first holds that trees that experience high temperatures more frequently during their youth, learn to buffer temperature-induced growth fluctuations. This allows them to survive
high temperature events without sacrificing much growth, which can be interpreted as a sign of higher fitness (DeSoto et al., 2020). This type of priming has been documented in the short-term: experiencing high heat triggers tolerance mechanisms, which can be more easily triggered in future (and potentially more damaging) high heat events, on a timescale of hours to months (Nievolia et al., 2017). Here we posit that priming could also work on a timescale of years to decades. The second hypothesis holds that trees that experience high temperatures more frequently during their youth, accumulate damage and lose the ability to respond optimally to high temperature events. This implies greater growth decreases under high temperature events during the rest of their life. A mechanism could be long-term damage to the tree’s root system for example.

Figure 1.3 illustrates our hypotheses using schematic temperature response curves. The first hypothesis, sketched on the left, implies that the temperature response curve of trees that experienced high temperatures more frequently early on, would shift right, as expressed by a higher temperature optimum. The slope measuring high temperature impacts could either become milder, stay the same, or become steeper—for example, if the new optimum gets closer to a local critical thermal limit. In any case, if we fix the breakpoint of the piecewise linear at the average (blue) temperature optimum, we expect the new (red) slope to become relatively milder. The second hypothesis, sketched on the right, implies that the temperature optimum would shift left or, if the temperature optimum is fixed at the average (blue) level, that the slope measuring high temperature impacts would become steeper, as trees accumulate damage and lose the ability to withstand high temperature events.

To test these hypotheses, we fix the breakpoint of the piecewise linear function to be the regional temperature optimum. The identification of the optimal breakpoint becomes imprecise in smaller samples. We prefer to rely on estimating changes in the impact of supraoptimal temperatures, under different degrees of early weather exposure. This is enough to distinguish the two hypotheses laid out above: hypothesis 1 corresponds to a milder impact of high temperatures, following more intense exposure to high temperatures during youth; hypothesis 2 corresponds to a stronger impact of high temperatures, following more intense exposure to high temperatures during youth. We consider the “priming” youth period for trees to be the first decade of their life, which we justify in appendix A.2.5. Our primary measure of early exposure to high temperatures is the annual number of high degree-days experienced by a tree, averaged over its youth. We also use alternative exposure measures. The model

\[4\] If trees acquire a greater tolerance to high temperatures, in response to a hotter youth, we could also expect them to acquire a greater tolerance to critically high temperatures, so that a higher critical thermal limit could apply to them. This likely depends on the severity of high temperatures experienced during youth.
Figure 1.3: **Acclimatization hypotheses.** The blue curve is the schematic temperature response curve of an average tree. The red curve is the schematic departure from this average response, for a tree that has experienced relatively more degree-days above the regional optimum. Left: acclimatization as “learning”. Right: acclimatization as accumulated damage.

can be laid out as follows:

\[
y_{ct} = \alpha_c + \beta_1 \text{low
degree-days}_{st} + \beta_2 \text{high
degree-days}_{st} \\
\quad + \gamma_{\text{lat,lon}} \ast \text{low
degree-days}_{st} + \beta_3 \text{youthExposure}_c \ast \text{low
degree-days}_{st} \\
\quad + \eta_{\text{lat,lon}} \ast \text{high
degree-days}_{st} + \beta_4 \text{youthExposure}_c \ast \text{high
degree-days}_{st} + \delta z_{ct} + \epsilon_{ct}
\]

where \( \text{youthExposure}_c = \frac{1}{10} \sum_{t=\text{year of birth}}^{10} \text{high
degree-days}_{st} \)

The subscript \( s \) denotes the weather grid cell to which tree core \( c \) belongs. The \( \text{youthExposure}_c \) variable has subscript \( c \) for tree core and not \( s \) for grid cell because the weather is averaged over a period that depends on when the tree was born. Threshold \( t \) is ecoregion-specific, and determined using the baseline model from the previous section. The high degree-day variable is interacted with location fixed effects \( \eta_{\text{lat,lon}} \), where lat and lon are rounded to the closest degree, so that the exposure interaction term captures differences in temperature impacts between trees that live in similar environments, but have experienced different weather histories early in life, due to differences in the timing of their birth. The term \( z_{ct} \) includes the same suite of controls as in the main model. We test the conflicting hypotheses \( \beta_4 > 0 \) (acclimatization as “learning”) versus \( \beta_4 < 0 \) (acclimatization as accumulated damage). We include a similar interaction term for the low degree-day variable.

In addition to the youth-adulthood timeline, we test for acclimatization on a decadal timescale. Specifically, we test whether the temperature response of a tree in a given decade, varies as a function
of the weather experienced during the previous decade. This allows for continual acclimatization to high temperatures throughout a tree’s life. The structure of the model is similar to that described above: high temperature exposure during a given decade, is interacted with high temperature exposure during the previous decade along with location fixed effects.

### 1.2.5.2 Adaptation analysis

We use our temperature model, to document adaptation to high temperatures. We first look for adaptation at the ecoregion level. We qualitatively compare the temperature response of tree growth in each US ecoregion, to its average climate. We then test adaptation more locally. We fix the optimal temperature breakpoint at its regional value\(^5\), and fit the corresponding piecewise linear model across 1°x1° grid cells. We regress the estimated upper slope value at each grid cell \(\hat{\beta}_{2_{\text{grid cell}}}\) (which captures the local impact of supraoptimal temperatures), on the average incidence of supraoptimal temperatures in each grid cell. We run this analysis both across large regions (humid ecoregions of the US, Cfa ecoregion), and within widely distributed species (Douglas Fir, Ponderosa Pine, and White Oak).

### 1.3 Results

#### 1.3.1 Temperature response curve

##### 1.3.1.1 Baseline model: response to year-round temperature

Figure 1.4 shows that in many ecoregions, temperature has a relatively neutral or slightly positive impact on tree growth up to a critical temperature threshold, beyond which tree growth decreases sharply with temperature. Regions Dfc and BSk are an exception: temperature initially has a mildly negative impact on tree growth, and a more negative impact beyond a critical threshold. The critical threshold varies by ecoregion: it is higher in humid regions (21°C in Dfc, 25°C in Dfb, 19°C in Dfa, and 26°C in Cfa) than in dry regions (12°C in Csb and 11°C in BSk). The negative impact of high temperatures, captured by the upper slope value, also varies in magnitude across ecoregions. Going from a 24-hour exposure to 26°C to a 24-hour exposure to 32°C, which corresponds to an additional 6 “high” degree days during the year, decreases annual tree ring width by 1.6% on average in Dfc, by 1.4% on average in Dfb, by 0.4% on average in Dfa and Csb, by 0.5% on average in BSk, and by 0.7% on average in Cfa.

\(^5\)Ideally, we would identify the locally-optimal threshold, but our threshold identification methodology is unreliable in small samples.
The more flexible spline-based curves are shown in appendix A.3.1. They suggest that a piecewise linear function is a good approximation of the growth-temperature relationship. We also test the sensitivity of the model to the search range for the optimal breakpoint. Restricting the search range to 5°-35°C yields very similar curves, although the temperature response is now exactly linear in BSk, up to a high optimal breakpoint situated in a region of coarsely estimated temperature impacts (see appendix A.3.2).

Figure 1.4: **Piecewise linear relationship between annual log ring width and temperature.** The optimal breakpoint is identified in the 1°-40°C range, using 10-fold cross-validation across years. The model includes tree fixed effects, a quadratic in total annual precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. The 95% confidence interval is added as a lighter area around the curve. Each graph corresponds to a different ecoregion of the US.

### 1.3.1.2 Seasonal model: response to spring-summer temperature

Figure 1.5 shows the spring-summer AMJJA temperature response curves by ecoregion. Most curves are very similar to the year-round curves shown above: the optimal breakpoint is slightly higher and the upper slope slightly steeper in ecoregions Dfb (27°C and -1.7% compared to 25°C and -1.4%)
and Csb (14°C and -0.6% compared to 12°C and -0.4%); the optimal breakpoint is slightly lower and
the upper slope slightly milder in ecoregions Dfa (16°C and -0.4% compared to 19°C and -0.4%) and
Cfa (25°C and -0.6% compared to 26°C and -0.7%). In BSk there is more of a change: the optimal
breakpoint is much lower at 4°C (compared to 11°C), and the negative impact of high temperatures is
stronger, it is an average decrease of 0.8% per 6 additional “high” degree days (compared to a 0.5%
decrease).

Figure 1.5: Piecewise linear relationship between annual log ring width and AMJJA tempera-
tenature. The optimal breakpoint is identified in the 1°-40°C range, using 10-fold cross-validation
across years. The model includes tree fixed effects, a quadratic in total annual precipitation, a linear
control in age, and standard errors clustered at the 2° gridcell level. The 95% confidence interval is
added as a lighter area around the curve. Each graph corresponds to a different ecoregion of the US.

1.3.1.3 Controlling for drought

Figure 1.6 shows the temperature response curves obtained for each ecoregion, after controlling for
drought using the 1-month, 3-month, and 12-month SPEI. All curves shift upwards slightly. In par-
ticular, we recover a more positive temperature effect at low temperatures in region BSk compared to
the baseline model, and the impact of high temperatures is less negative in all ecoregions, except in region BSk. The optimal breakpoint does not change much: it shifts up by two degrees in BSk and Cfa, shifts down by one degree in Dfc and Dfb, does not change in Csb. It does decrease from 19°C to 9°C in Dfa, although the overall curve is very flat.

Figure 1.7 shows the spring-summer temperature response curves, after controlling for drought. All curves shift up and right, compared to baseline AMJJA curves. Overall, the impact of low temperatures is more positive and the optimal breakpoint lower in dry ecoregions (Csb and BSk) than in humid ecoregions (Dfa, Dfb, Dfc, and Cfa). In appendix A.3.3, a side-by-side display of the year-round and spring-summer responses, with and without drought controls, for each ecoregion, makes those comparisons easier.

Figure 1.6: Piecewise linear relationship between annual log ring width and temperature, controlling for drought. The optimal breakpoint is identified in the 1°-40°C range, using 10-fold cross-validation across years. The model includes tree fixed effects, monthly values of the SPEI 1, 3, and 12, a quadratic in total annual precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. The 95% confidence interval is added as a lighter area around the curve. Each graph corresponds to a different ecoregion of the US.
Figure 1.7: **Piecewise linear relationship between annual log ring width and AMJJA temperature, controlling for drought.** The optimal breakpoint is identified in the 1°-40°C range, using 10-fold cross-validation across years. The model includes tree fixed effects, monthly values of the SPEI 1, 3, and 12, a quadratic in total annual precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. The 95% confidence interval is added as a lighter area around the curve. Each graph corresponds to a different ecoregion of the US.

1.3.1.4 **Sample restrictions: temperature-sensitive trees, and isohydric versus anisohydric trees**

Figure 1.8 shows the spring-summer temperature response curve of temperature-sensitive trees. Increases in temperature have a strong positive impact on tree growth until the optimal level of 32°C, beyond which the impact turns negative. In practice, an additional 6 degree days below 32°C translate into a 0.2% increase in tree ring width on average, and an additional 6 degree days above 32°C translate into a 2.8% decrease in tree ring width on average. This curve is closer to the leaf-level response of net photosynthesis to temperature, than the previous curves. It is not impacted by the inclusion of drought into the model (see appendix A.3.5), consistent with temperature being the limiting growth factor for this sample of trees.
Figure 1.8: **Piecewise linear relationship between annual log ring width and AMJJA temperature for temperature-sensitive trees.** The optimal breakpoint is identified in the 1°-40°C range, using 10-fold cross-validation across years. The model includes tree fixed effects, a quadratic in total annual precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. The 95% confidence interval is added as a lighter area around the curve. The sample is restricted to temperature-sensitive trees.

Temperature curves by isohydric status are shown in appendix A.3.6. Across the three ecoregions where isohydric and anisohydric species overlap, isohydric trees seem to suffer less from high temperatures, which is consistent with their greater stomatal control: the negative impact of supraoptimal temperatures is milder for isohydric species. The optimal breakpoint is higher for anisohydric species in regions Cfa and Csb, which could be consistent with their riskier strategy (they can afford a riskier strategy if they can tolerate a greater range of temperatures). However, the pattern is the reverse in region Dfb, and we cannot interpret the difference in optimal breakpoints as being statistically significant.
1.3.1.5 Lagged models

Figure 1.9 shows the temperature response curves by ecoregion, after controlling for a one-year lag in temperature and precipitation. A positive impact of low temperatures appears in dry ecoregions Csb and BSk. All optimal breakpoints increase slightly compared to baseline curves, except in Dfb (slight decrease from 25°C to 23°C) and Dfa (decrease from 19°C to 14°C, although the overall curve is very flat). The negative impact of high temperatures is milder in all ecoregions, compared to the baseline model.

Figure 1.9: Piecewise linear relationship between annual log ring width and temperature, **controlling for one lag in weather**. The optimal breakpoint is identified in the 1°-40°C range, using 10-fold cross-validation across years. The model includes tree fixed effects, a quadratic in total annual precipitation, a 1-year lag in temperature and precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. The 95% confidence interval is added as a lighter area around the curve. Each graph corresponds to a different ecoregion of the US.

High temperature events could have a lagged impact on tree radial growth, either by reducing growth over multiple years, or if tree growth tends to “rebound” to compensate for low-growth years. The cumulative impact of high temperatures over the long term could thus be negative or neutral. We
use a distributed lag model with 10 lags, to document whether temperature has a persistent effect on tree radial growth over time. The coefficients on all lags of low degree days and high degree days are shown in appendix A.3.7. Figure A.16, which corresponds to a year-round temperature model, shows that the cumulative impact of high temperatures is negative in all ecoregions. There is evidence of negative lagged impacts in ecoregions Dfc, Dfb, and Csb, and rebound growth in ecoregion Dfa. The magnitude of the contemporaneous impact of high temperatures is always greater than the magnitude of lagged impacts. Ecoregion Cfa is notable for its clear absence of lagged impacts.

Figure A.17, which corresponds to an AMJJA temperature model, shows a very similar pattern. Only ecoregion BSk looks quite different, which might be because a very low threshold of 4°C is used in this model. Figure A.18, which corresponds to an AMJJA temperature model for temperature-sensitive trees, shows a similar pattern to that of ecoregion Cfa: high temperatures do not have lagged impacts on the growth of this sample of trees.

1.3.2 Applications

1.3.2.1 Acclimatization to high temperatures

Regression tables for the acclimatization model are shown in appendix A.3.8. Figure 1.10 highlights the value of the interaction between high degree-days and early weather exposure. A positive value indicates acclimatization as “learning”, and a negative value indicates acclimatization as accumulated damage. The three rows in each graph correspond to three variations of the model, which use different exposure variables: starting from the annual sum of high degree-days experienced by a tree, (1) “average exposure” is the average value of this sum over the tree’s youth; (2) “maximum exposure” is the maximum value over the tree’s youth; and (3) “90th percentile exposure” is the 90th percentile over the tree’s youth.

This interaction coefficient is very small across models and quadrants, even when it is significantly different from zero, like in region Dfa. This suggests that there is no acclimatization to high temperatures at this timescale. Replacing 1° location fixed effects, which are interacted with high degree days, by a measure of the average climate at each location, to control for differences in the baseline environment while absorbing less variation, does not change the results. Similar acclimatization hypotheses are tested for tree growth response to AMJJA temperatures. The results, which are included in appendix A.3.9, also point to the absence of acclimatization at this time scale.

We then test the existence of acclimatization to JJA temperatures among temperature-sensitive
Figure 1.10: **Main coefficient estimates from an ecoregion-level acclimatization model.** The optimal temperature breakpoint used to define low and high degree days is fixed at the ecoregion baseline optimum. The model also includes tree fixed effects, 1° location fixed effects interacted with high degree days, a quadratic in total annual precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. Three variations of the model are estimated, which correspond to three different measures of early tree exposure to high temperatures.

trees. We find some evidence of acclimatization as accumulated damage, illustrated in Figure 1.11, with corresponding regression tables in appendix A.3.10. Based on the model that uses the first exposure metric avg \( d_{dayHigh} \), the magnitude of acclimatization is as follows: for a 1 standard deviation increase in the frequency of high temperatures during youth—which corresponds to an extra 8.6 degree-days above 32°C experienced annually during the first 10 years of life—the harmful impact of high temperatures later in life increases by 9% or 0.07 percentage points.

We also explore a decadal acclimatization timescale, where trees adapt their temperature response from one decade to the next, based on weather experienced during the previous decade. Regression tables for the decadal acclimatization model are shown in appendix A.3.11. We find evidence of positive acclimatization in Dfa and negative acclimatization in Dfc and Cfa, but magnitudes are as small as in the longer youth-to-adulthood timescale.

### 1.3.2.2 Adaptation to high temperatures

The temperature response curves documented in the previous sections reveal interesting cross-ecoregion patterns. First, the optimal temperature breakpoint is lower in dry regions than in humid regions. This is consistent with the temperature-mediated drought channel being more prevalent in dry re-
Figure 1.11: **Main coefficient estimates from a US-wide acclimatization model, restricted to the sample of temperature-sensitive trees.** The optimal temperature breakpoint used to define low and high degree days is fixed at the baseline optimum for this sample of trees. The model also includes tree fixed effects, 1° location fixed effects interacted with high degree days, a quadratic in total annual precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. Three variations of the model are estimated, which correspond to three different measures of early tree exposure to high temperatures.

Second, trees in ecoregions characterized with warmer summers tend to suffer less from high temperatures: the impact of supraoptimal temperatures in lower in Dfa (hot summer) compared to Dfb (warm summer) and lower in Dfb compared to Dfc (cool summer) in all but one specification. This suggests that trees are better able to withstand high temperatures in climates characterized by more frequent high temperatures.

We further explore adaptation by testing whether the harmful impact of supraoptimal temperatures varies with the local climate. We start with all humid ecoregions combined, excluding dry ecoregions because their optimal temperature breakpoint is lower than in humid ecoregions. We find that the optimal temperature in this combined region is 23°C, based on year-round temperatures. On the left, figure 1.12 shows the average incidence of degree-days above 23°C, by 1° grid cell. On the right, it shows the harmful impact on annual tree growth of an additional degree-day above 23°C, by 1° grid cell. It only includes grid cells where the impact of high degree-days is negative and statistically significant. The full map is reported in appendix A.3.12. We see that trees in the West seem to

---

6The significance threshold we use is a p-value of the projected F statistic (excluding fixed effects) of 0.05.
suffer relatively more from supraoptimal temperatures, consistent with the lower incidence of such
temperatures. There also seems to be a slight gradient in the East, where trees suffer slightly more
from supraoptimal temperatures at more Northern latitudes, although the pattern is not nearly as
stark as the gradient in high temperature incidence observed on the left map. We document this
relationship more formally by regressing the impact of high degree-days on the average incidence of
degree-days. As shown on figure 1.13, we find a positive and statistically significant coefficient: at the
average level of high temperature incidence, the impact of six additional degree-days above 23°C is a
decrease in tree growth of 0.96%; one standard deviation above the average incidence level, the impact
of six additional degree-days above 23°C is a decrease in tree growth of 0.55%. This relationship is
mostly driven by the lower tolerance to high temperature in locations where high temperatures are
very rare.

Figure 1.12: Adaptation maps: frequency and impact of degree-days above 23°C in humid
ecoregions (selected grid cells). [Left] Average number of annual degree-days above 23°C, by 1°
grid cell. [Right] Impact on annual tree growth of an additional degree-day above 23°C, in humid
ecoregions. The model is run separately on each 1° grid cell. Only negative and statistically significant
coefficients are displayed, for graphical clarity.

We repeat this methodology on different regions and within the most widely distributed species: all
trees in the south-eastern Cfa ecoregion (year-round optimal temperature of 26°C), Douglas Fir trees
(year-round optimal temperature of 17°C), Ponderosa Pine trees (year-round optimal temperature of
14°C), and White Oak trees (year-round optimal temperature of 23°C). Although maps suggest that
some adaptation might be going on (see appendix A.3.13), regression models find a non-statistically
significant correlation between the average incidence of high temperatures and the impact of high
temperatures. This association is statistically significant only in the Ponderosa Pine sample: it is
positive, and of a smaller magnitude compared to the case of all humid ecoregions. The negative impact of supraoptimal temperatures goes from -0.64% to -0.51% among Ponderosa Pine trees, as the average incidence of supraoptimal temperatures increases by one standard deviation.

1.4 Discussion

Trees display a clear non-linear response to temperature, based on their annual growth rings. In particular, high temperatures have a strong negative impact on current tree growth. Six additional degree-days above the local optimal temperature level lead to an average decrease in annual tree ring width of 1%, with variations across US ecoregions. This impact is not compensated for by rebound growth in the following years, contrary to what has been observed for extreme drought events (DeSoto et al., 2020), except in region Dfa where some rebound growth is visible. In most ecoregions, supraoptimal temperatures continue to have a negative impact on growth over a couple years. However, the concurrent impact of high temperatures on tree growth is stronger than lagged impacts. This is also what George (2014) observes for the impact of average summer temperature.

The key breakpoint where the non-linearity kicks in also varies across ecoregions. It is about 20-
25°C in humid ecoregions and 10-15°C in dry ecoregions, if year-round temperatures are considered. These breakpoints are colder than the temperature optima suggested by experimental evidence on the thermal performance of net photosynthesis. However, in humid ecoregions, our optimal breakpoints are on par with the optimal temperature levels derived at the ecosystem-level, which vary between 15°C and 30 °C across the US, and combine direct and indirect temperature effects like our estimates do. Dry ecoregions, BSk and Csb, still display a more puzzling lower breakpoint, with a nearly linearly decreasing curve for BSk.

At least three factors could help explain this puzzling pattern. First, in regions where temperatures overlap across seasons, or in regions with multiple strong seasonal responses, the year-round temperature response curves could be masking seasonal heterogeneity. In particular, regions BSk and Csb are characterized by a more condensed and nearly unimodal temperature distribution, compared to the bimodal temperature distributions of the other ecoregions. Second, our estimates reflect both direct temperature impacts on carbon exchange processes, and indirect temperature impacts through factors that covary with temperature over time and impact tree radial growth. One such factor is likely to be drought, since the ITRDB dataset is biased towards drought-sensitive trees. Since drought suppresses tree growth, temperature-mediated drought impacts likely pull the temperature response curves downwards. Third, weather likely has a lagged impact on tree growth, especially recent weather, and there is likely temporal autocorrelation in temperature, so our temperature curves could partially reflect lagged temperature impacts.

Both AMJJA temperature response curves that control for drought, and year-round response curves that control for lagged weather—likely correlated with current drought—yield more intuitive curves. The impact of suboptimal temperatures becomes more positive, the negative impact of supraoptimal temperatures becomes milder, and optimal breakpoints increase. These breakpoints are close to 15°C in both BSk and Csb. These optimal levels still seem low if response curves purely reflect direct temperature impacts. Given that drought is a multi-facted phenomenon, and that we do not include data on soil moisture, it could be that our temperature response curves are still impacted by leftover variation in temperature-mediated drought conditions, after controlling for current drought or past weather. There could also be other factors, like nutrient availability or pest pressure, that influence our temperature response curves, if those are affected by medium to high temperatures.

We connect our temperature response estimates with the physiological literature by focusing on two particular samples. First, restricting the data to temperature-sensitive trees yields a curve, which is close to what would be predicted by extrapolating experimental evidence on direct temperature
effects. Specifically, spring and summer temperatures have a positive impact on tree growth up until 32°C, above which further temperature increases are harmful for growth. Second, we find that the negative impact of supraoptimal temperatures is milder for isohydric species compared to anisohydric species, which is consistent with their more conservative management of stomatal opening.

Our methodology for identifying temperature response curves requires large samples and enough variation to overcome tree core fixed effects. Our dataset is very large in terms of observations but less so in terms of locations, whereas temperature only varies by location. This limits us for exploring heterogeneity in trees’ temperature response, beyond ecoregions, temperature-sensitive trees, and isohydric vs anisohydric species. For instance, we cannot reliably estimate temperature response curves across species within ecoregions, or by elevation in mountainous ecoregions.

Using our new temperature model, we document whether weather experienced during a tree’s youth shapes this tree’s later response to temperature, relative the the average local response\(^7\). We do not find evidence of acclimatization at this timescale, except among temperature-sensitive trees. The latter fit our negative acclimatization hypothesis, whereby experiencing high temperatures more often during youth leads a tree to suffer more from high temperatures during the rest of its life. We also explore a decadal acclimatization timescale, where trees adapt their temperature response from one decade to the next, based on weather experienced during the previous decade. We find evidence of positive acclimatization in Dfa and negative acclimatization in Cfa and Dfc to some extent, but magnitudes are very small. Trees that have a higher temperature optimum, could be operating closer to the end of their safe range, compared to trees that have a lower temperature optimum (O’sullivan et al., 2017). Supraoptimal temperatures could thus be more harmful events for trees with a higher temperature optimum, which could explain the negative acclimatization among temperature-sensitive trees and in region Cfa.

One thing to note is that our acclimatization results could reflect seedling or sapling selection, in addition or instead of phenotypic adaptation. For example, high temperatures could lead to higher seedling mortality hence less tree births, or to greater selection of temperature-resistant seedlings. This is difficult to test rigorously since the date of the first ring is different from the “birth” date of the tree. As a sanity check, we verify that the number of births per location and year in our dataset is not statistically significantly correlated with high temperature exposure.

We also document whether the average climate influences a tree’s response to temperature, and

\(^7\)This detail is important. We exploit temporal variations in weather around the local average climate, and not spatial variation in weather across locations with different climates.
more specifically its response to high temperatures. We find some evidence of adaptation to high tem-
peratures in humid ecoregions, in the sense that trees are less harmed by supraoptimal temperatures,
in locations where such temperatures are more frequent. High temperatures are 43\% less harmful to
tree growth, when the average incidence of high temperatures increases by one standard deviation.
This seems to be driven by differences between locations that experience very different frequencies of
high temperatures. We do not find much evidence of adaptation in regions or samples with smaller
climatic gradients.

Our real-life, tree-level temperature response curves hold a lot of promise for shedding light on
the harmful impact of high temperatures, on tree growth projections under expected shifts in the
temperature distribution, and on the potential for acclimatization and long-run adaptation. They
could be used for example to shed light on the divergence problem at northern latitudes, where
decreased sensitivity of tree growth to average temperature has been observed, and could be due
to recent shifts in the temperature distribution, combined with a non linear response of tree growth
to temperature (D’Arrigo et al., 2008). The scope of our results is limited by strong sampling bias in
the ITRDB data, but we can reproduce our methodology on more representative data like tree ring
measurements from the Forest Inventory and Analysis dataset, if those are made accessible.
Appendix A

Appendix to Chapter 1
A.1 Additional data

A.1.1 Individual maps of Köppen-Geiger ecoregions

Figure A.1: Location of tree ring sequences, by Köppen-Geiger ecoregion. Circle size indicates the number of tree cores sampled at each location.
A.1.2 Map of dominant species

Figure A.2: **Location of tree ring sequences, by species.** Circle size indicates the number of tree cores sampled at each location, color indicates tree species. The sample is restricted to the most represented species in the ITRDB data. PIEC: Shortleaf Pine, PIED: Pinyon Pine, PIPO: Ponderosa Pine, PSME: Douglas Fir, QUAL: White Oak, QUDG: Blue Oak, QUST: Post Oak, TADI: Bald Cypress, TSCA: Eastern Hemlock, TSME: Mountain Hemlock.
A.1.3 Map of isohydric and anisohydric species

Figure A.3: Location of tree ring sequences, by isohydric status. Circle size indicates the number of tree cores sampled at each location.
A.2 Additional methods

A.2.1 Tree growth data: formatting

We downloaded all tree ring records from the International Tree-Ring Data Bank (ITRDB), using the National Oceanic and Atmospheric Administration (NOAA) API. We filtered records geographically, to cover the contiguous US. Raw measurement files in the ITRDB are available in two different formats. The Tucson decadal format starts with a 3-line metadata header, and each tree core sequence is broken down across lines by decade. The NOAA format includes a more detailed metadata header, and reports the data in matrix form, with rows denoting years and columns denoting tree core identifiers. Files in Tucson format have less formatting issues, especially in the core identifier labels, so we extract measurements from files in Tucson format and match them with metadata extracted from the corresponding files in NOAA format. Section A.2.2 describes the list of corrections that we made to the raw data. We corrected multiple formatting issues, including: typographic errors in decades, tree core identifiers, or end-of-series markers that indicate the unit of measurement; extra characters; missing headers; trailing zeros at the end of a sequence; multiplicity of codes used for missing values. We also flagged and dropped duplicate sequences within and across files, whether those were reported under the same tree core identifier or under different tree core identifiers. We addressed duplicate use of tree core identifiers by creating new unique identifiers. We compared our list of formatting corrections with that of Zhao et al. (2019), who performed a similar exercise. Our list is at least as comprehensive as their’s. In addition, we flagged and corrected suspicious ring width measurements. For example, measurements are either reported in units of 0.01mm or 0.001mm, which is indicated by two different end-of-series markers. Some markers are clearly erroneous (e.g. all sequences but one are reported under the same marker, while the distinct sequence is of the same order of magnitude as the others, which results in oddly high measurements). Flagging measurements that exceed 30 times the median of the sequence revealed further issues that were addressed and reported in Section A.2.2. This threshold of 30 was determined by exploring various thresholds in decreasing order, until the flagged measurements no longer suggested clear typographical errors.

1This description of the Tucson decadal format highlights some peculiarities encountered in the data, including the issue of various codes for missing values and trailing zeros. [link](http://www.cybis.se/wiki/index.php?title=Tucson_format#cite_note-0)

2Sometimes, the same identifier is used for two non-duplicate core sequences in a file. Cases vary between consecutive and non-consecutive sequences, years overlapping or not across the two sequences, measurements being of the same or of a different order of magnitude. In all cases, we consider that the sequences come from two different cores, and assign them a new unique identifier.
## A.2.2 Tree growth data: list of corrections

<table>
<thead>
<tr>
<th>Action</th>
<th>Automated</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deleting fully repeated sequences</td>
<td>Yes</td>
<td>Deletion of sequences that are identical with only minor differences.</td>
</tr>
<tr>
<td>Renaming consecutive sequences that have the same tree ID and different measurements</td>
<td>Yes</td>
<td>Ensure unique identifier for each sequence.</td>
</tr>
<tr>
<td>Deleting missing years</td>
<td>Yes</td>
<td>Ensure complete data series.</td>
</tr>
<tr>
<td>Removing duplicated part of 32-13C sequence</td>
<td>Yes</td>
<td>Remove redundant data.</td>
</tr>
<tr>
<td>Replacing decade labels of line 324, 325 with 20**</td>
<td>Yes</td>
<td>Correct date notation.</td>
</tr>
<tr>
<td>Replacing core labels of line 160 - 168 (HH09B') with HH09B</td>
<td>Yes</td>
<td>Ensure consistent labeling.</td>
</tr>
<tr>
<td>Inserting # before the site name</td>
<td>Yes</td>
<td>Standardize naming convention.</td>
</tr>
<tr>
<td>Deleting tree &quot;WNCH15-1&quot; (odd years)</td>
<td>Yes</td>
<td>Remove unnecessary data.</td>
</tr>
<tr>
<td>Renaming (not removing) duplicated WLD35B</td>
<td>Yes</td>
<td>Keep data for analysis.</td>
</tr>
<tr>
<td>Replacing decade 1924 with 1923 (tree 754013), 1935 by 1930 (tree 628071), 1925 by 1924 (tree 607023), 1887 by 1886 (tree 538042), 1815 by 1810 (tree 369049), 1857 by 1855 (tree 224005), 1889 by 1888 (tre 215049), 1877 by 1876 (tree 115025)</td>
<td>Yes</td>
<td>Adjust data for accuracy.</td>
</tr>
<tr>
<td>Removing line 1126, 2242, 2680</td>
<td>Yes</td>
<td>Eliminate invalid data.</td>
</tr>
<tr>
<td>Renaming (not removing) duplicated lohc02</td>
<td>Yes</td>
<td>Keep track for analysis.</td>
</tr>
<tr>
<td>Renaming duplicated ccc063, ccc072</td>
<td>Yes</td>
<td>Ensure unique identifiers.</td>
</tr>
<tr>
<td>Renaming (not removing) duplicated prd221</td>
<td>Yes</td>
<td>Keep data for analysis.</td>
</tr>
<tr>
<td>Renaming (not removing) duplicated wm4-3 (not WM2-1a)</td>
<td>Yes</td>
<td>Ensure unique identifiers.</td>
</tr>
<tr>
<td>Replacing the decade label of line 80 (&amp;) with 10</td>
<td>Yes</td>
<td>Correct date notation.</td>
</tr>
<tr>
<td>Deleting measurement of tree WTK08B year 1933</td>
<td>Yes</td>
<td>Remove invalid data.</td>
</tr>
<tr>
<td>Replacing measurement of tree SUN09B year 1665 (17996) by 1796</td>
<td>Yes</td>
<td>Correct measurement.</td>
</tr>
<tr>
<td>Replacing measurement of tree 501022 year 1926 (4300) by 43; inserting a zero measurement right after and shifting the next two measurements by one year</td>
<td>Yes</td>
<td>Ensure accurate data series.</td>
</tr>
<tr>
<td>Automatically flagged: ratio of width to median width exceeds 30</td>
<td>Yes</td>
<td>Identify potential data anomalies.</td>
</tr>
</tbody>
</table>

### Key Points
- Automated procedures are applied to streamline data processing.
- Manual checks and interventions are applied for specific corrections.
- Data integrity and accuracy are ensured through systematic verification and validation.
A.2.3 Temperature response curve: age control and standard errors

The age control allows for a linearly decreasing growth rate over a tree’s lifetime. It is likely not as flexible as the diversity and complexity of tree age growth curves would call for, but in practice many trees exhibit a pattern of log ring width decreasing linearly with age (Schofield et al., 2016). In addition, the slope of that relationship likely varies over species and locations, so we test and confirm the robustness of our estimates to allowing for varying age slopes. This age control is not necessary if tree age and temperature are not correlated. However there is both a slight trend upwards in age over time and a slight trend upwards in average and maximum temperature, hence the inclusion of the control.

The choice of $2^\circ$ reflects the fact that PRISM temperature interpolation is based on about 1800 stations, i.e. approximately 1 station every 0.8 degrees across the US, and the seven closest stations inform the temperature estimate at any given location so a relevant correlation distance would be between $1.5^\circ$ and $2.5^\circ$. In addition, errors in ring width measurement and crossdating are likely to be correlated across cores of a given tree, across trees at the same sampling site or by researcher. This could introduce temporal correlation in the errors, so we alternatively use Conley-West standard errors. The Conley-West approach allows for the error terms to be correlated spatially and temporally, up to chosen thresholds of 250km and 10-years, so that the precision of the estimates is not inflated. We do not use this method as the default because it is very demanding computationally, and does not affect our main conclusions.

A.2.4 Temperature response curve: spline model

To allow for a flexible response of annual tree growth to temperature, we use a restricted cubic spline in temperature with 8 knots (at 0, 5, 10, 15, 20, 25, 30, and 35°C), which results in $N = 7$ temperature variables. Essentially, separate cubic polynomials are fit on each temperature segment, where segments are defined by the position of the knots. They are fit in a way that each segment curve smoothly joins with the others at the knots. In addition, the restricted cubic spline imposes that the curve be linear before the first knot and after the last knot. This generally provides a better fit to the data, and avoids poor behavior at the tails. The spline model exploits the entire temporal distribution of temperature over a year. The spline transformation is applied to the half-hourly temperature data, and the resulting spline variables are summed by year and scaled by 48 to get a daily interpretation. As with the degree-day structure of the main model the key assumption is that temperature effects on tree growth are
cumulative over time and annual tree growth is proportional to total exposure.

A.2.5 Acclimatization analysis: choice of age cutoff characterizing tree youth

We define young trees as trees below 10 years of age, with age defined as years since the first ring measurement in a tree core sequence. We assume that this metric is a good proxy for years since the formation of the pith. The 10 year cutoff is chosen based on growth fluctuations observed in the data. We compare ring width variability across age decades, after controlling for differences in fluctuations due to age-driven differences in growth levels (we transform ring width into the ratio relative to a tree’s decadal mean), and controlling for baseline differences in growth fluctuations across trees (we scale our transformed ring width metric using each tree’s lifelong standard deviation in growth). The leftover differences in growth variability across age decades, shown in Figure A.4, suggest that growth fluctuates slightly more in the first decade, relative to subsequent decades. The most recent decades are excluded from the comparison due to small sample sizes, reported in histogram form at the top of the graph.

Figure A.4: Variation in tree growth as a function of tree age. Ring width measurements are centered and scaled at the tree core and age decade level, standardized using tree core level standard deviation, and converted to absolute deviations. The histogram on top reflects the frequency of tree ring measurements corresponding to each age decade.
A.3 Additional results

A.3.1 Temperature response curve: spline model

Fluctuations are generally influenced by the position of the spline knots, but this seems particularly true at high temperatures, where sharp increases appear in ecoregions Csb and BSk, right at the before-last spline knot of 30°C. These increasing tails do not survive the transition to a piecewise linear function, even when imposing a relatively high breakpoint. This suggests that splines are not adapted to estimating high temperature impacts in our data.
A.3.2 Temperature response curve: restricted range for optimal breakpoint search

Figure A.6: Piecewise linear relationship between annual log ring width and temperature, restricted search range for the breakpoint. The optimal breakpoint is identified in the 5°-35°C range, using 10-fold cross-validation across years. The model includes tree fixed effects, a quadratic in total annual precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. The 95% confidence interval is added as a lighter area around the curve. Each graph corresponds to a different ecoregion of the US.
A.3.3 Temperature response curve, map, and species composition, by ecoregion

(a) Dfc, locations

(b) Dfc, tree species

(c) Dfc, year-round response without drought control

(d) Dfc, year-round response with drought control

(e) Dfc, AMJJA response without drought control

(f) Dfc, AMJJA response with drought control
(a) Dfb, locations

(b) Dfb, tree species

(c) Dfb, year-round response without drought control

(d) Dfb, year-round response with drought control

(e) Dfb, AMJJA response without drought control

(f) Dfb, AMJJA response with drought control
(a) Csb, locations

(b) Csb, tree species

(c) Csb, year-round response without drought control

(d) Csb, year-round response with drought control

(e) Csb, AMJJA response without drought control

(f) Csb, AMJJA response with drought control
(a) BSk, locations

(b) BSk, tree species

(c) BSk, year-round response without drought control

(d) BSk, year-round response with drought control

(e) BSk, AMJJA response without drought control

(f) BSk, AMJJA response with drought control
A.3.4 Map of temperature-sensitive trees

Figure A.13: Location of tree ring sequences and frequency of measurements over time. Circle size indicates the number of tree cores sampled at each location. The sample is restricted to temperature-sensitive trees in the ITRDB data.
A.3.5 Temperature response curve: temperature-sensitive trees, controlling for drought

Figure A.14: Piecewise linear relationship between annual log ring width and temperature for temperature-sensitive trees, controlling for drought. The optimal breakpoint is identified in the 1°-40°C range, using 10-fold cross-validation across years. The model includes tree fixed effects, monthly values of the SPEI 1, 3, and 12, a quadratic in total annual precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. The 95% confidence interval is added as a lighter area around the curve. The sample is restricted to temperature-sensitive trees.
A.3.6 Temperature response curve: splitting Eastern ecoregions by isohydric status

Figure A.15: Piecewise linear relationship between annual log ring width and temperature, by isohydric status. The optimal breakpoint is identified in the 1°-40°C range, using 10-fold cross-validation across years. The model includes tree fixed effects, a quadratic in total annual precipitation, a linear control in age, and standard errors clustered at the 2° gridcell level. Graphs are split by ecoregion (top to bottom) and by isohydric status (anisohydric on the left, isohydric on the right).
A.3.7 Legacy effects of temperature: distributed lag model

Figure A.16: Temperature coefficients from a distributed lag model. The outcome is annual log ring width, and the model includes 10 lags in low degree-days and high degree-days. The temperature threshold used to distinguish low and high degree-days is the optimal breakpoint identified in the baseline model for each ecoregion. The model also includes tree fixed effects, a quadratic in total annual precipitation along with 10 lags, a linear control in age, and standard errors clustered at the 2° gridcell level. Each graph corresponds to a different ecoregion of the US.
Figure A.17: **AMJJA temperature coefficients from a distributed lag model.** The outcome is annual log ring width, and the model includes 10 lags in low degree-days and high degree-days. The temperature threshold used to distinguish low and high degree-days is the optimal breakpoint identified in the baseline AMJJA model for each ecoregion. The model also includes tree fixed effects, a quadratic in total annual precipitation along with 10 lags, a linear control in age, and standard errors clustered at the 2° gridcell level. Each graph corresponds to a different ecoregion of the US.
Figure A.18: **AMJJA temperature coefficients from a distributed lag model for temperature-sensitive trees.** The outcome is annual log ring width, and the model includes 10 lags in low degree-days and high degree-days. The temperature threshold used to distinguish low and high degree-days is the optimal breakpoint identified in the baseline AMJJA model for temperature-sensitive trees. The model also includes tree fixed effects, a quadratic in total annual precipitation along with 10 lags, a linear control in age, and standard errors clustered at the 2° gridcell level. The sample is restricted to temperature-sensitive trees.

### A.3.8 Acclimatization: year-round temperature response

The following tables show estimates of the main coefficients from the acclimatization model: **low ddays** is the impact of degree-days below the optimal breakpoint, **low ddays X exposure** captures the change in the impact of degree-days below the optimal breakpoint, as early exposure to high degree days increases, **high ddays** is the impact of degree-days above the optimal breakpoint, and **high ddays X exposure** captures the change in the impact of degree-days above the optimal breakpoint, as early exposure to high degree days increases. Each column corresponds to a different measure of early weather exposure: starting from the annual sum of high degree-days experienced by a tree, (1) **avg ddayHigh** is the average value of this sum over the tree’s youth; (2) **max ddayHigh** is the maximum
value over the tree’s youth; and (3) $p90\ ddayHigh$ is the 90th percentile over the tree’s youth.

Table A.1: Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1) avg ddayHigh</th>
<th>(2) max ddayHigh</th>
<th>(3) p90 ddayHigh</th>
</tr>
</thead>
<tbody>
<tr>
<td>low ddays</td>
<td>-0.000667**</td>
<td>-0.000851**</td>
<td>-0.000893**</td>
</tr>
<tr>
<td></td>
<td>(0.000209)</td>
<td>(0.000282)</td>
<td>(0.000253)</td>
</tr>
<tr>
<td>low ddays X exposure</td>
<td>0.0000130*</td>
<td>0.0000150*</td>
<td>0.0000173**</td>
</tr>
<tr>
<td></td>
<td>(0.00000552)</td>
<td>(0.00000619)</td>
<td>(0.00000548)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.0103***</td>
<td>-0.00825*</td>
<td>-0.00848**</td>
</tr>
<tr>
<td></td>
<td>(0.00181)</td>
<td>(0.00372)</td>
<td>(0.00329)</td>
</tr>
<tr>
<td>high ddays X exposure</td>
<td>-0.0000706</td>
<td>-0.000119</td>
<td>-0.000117</td>
</tr>
<tr>
<td></td>
<td>(0.00000501)</td>
<td>(0.000100)</td>
<td>(0.0000899)</td>
</tr>
<tr>
<td>Observations</td>
<td>4330</td>
<td>4330</td>
<td>4330</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.590</td>
<td>0.590</td>
<td>0.590</td>
</tr>
</tbody>
</table>


Table A.2: Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1) avg ddayHigh</th>
<th>(2) max ddayHigh</th>
<th>(3) p90 ddayHigh</th>
</tr>
</thead>
<tbody>
<tr>
<td>low ddays</td>
<td>-0.00148***</td>
<td>-0.00150***</td>
<td>-0.00146***</td>
</tr>
<tr>
<td></td>
<td>(0.0000758)</td>
<td>(0.0000648)</td>
<td>(0.0000820)</td>
</tr>
<tr>
<td>low ddays X exposure</td>
<td>-0.00000386</td>
<td>0.000000290</td>
<td>-0.000000290</td>
</tr>
<tr>
<td></td>
<td>(0.00000325)</td>
<td>(0.00000158)</td>
<td>(0.00000242)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00582***</td>
<td>-0.00579***</td>
<td>-0.00574***</td>
</tr>
<tr>
<td></td>
<td>(0.0000715)</td>
<td>(0.00000618)</td>
<td>(0.0000751)</td>
</tr>
<tr>
<td>high ddays X exposure</td>
<td>0.000000296</td>
<td>0.000000472</td>
<td>0.000000141</td>
</tr>
<tr>
<td></td>
<td>(0.00000107)</td>
<td>(0.000000622)</td>
<td>(0.00000106)</td>
</tr>
<tr>
<td>Observations</td>
<td>102207</td>
<td>102207</td>
<td>102207</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.648</td>
<td>0.648</td>
<td>0.648</td>
</tr>
</tbody>
</table>

Ecoregion: Dfb, temperature threshold: 25.
Table A.3: Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg ddayHigh</td>
<td>0.000890*</td>
<td>0.000357</td>
<td>0.000539*</td>
</tr>
<tr>
<td></td>
<td>(0.000473)</td>
<td>(0.000260)</td>
<td>(0.000279)</td>
</tr>
<tr>
<td>max ddayHigh</td>
<td>0.000890</td>
<td>0.000615</td>
<td>0.000611</td>
</tr>
<tr>
<td></td>
<td>(0.000473)</td>
<td>(0.000260)</td>
<td>(0.000279)</td>
</tr>
<tr>
<td>p90 ddayHigh</td>
<td>0.000890</td>
<td>0.000615</td>
<td>0.000611</td>
</tr>
<tr>
<td></td>
<td>(0.000473)</td>
<td>(0.000260)</td>
<td>(0.000279)</td>
</tr>
</tbody>
</table>

low ddays X exposure

|                | -0.00000105          | -0.000000284         | -0.000000522         |
|                | (0.000000655)        | (0.000000306)        | (0.000000358)        |
| high ddays     | -0.00272**           | -0.00275***          | -0.00241***          |
|                | (0.00106)            | (0.000485)           | (0.000433)           |
| high ddays     | -0.00272             | -0.00275             | -0.00241             |
|                | (0.00106)            | (0.000485)           | (0.000433)           |
| high ddays X exposure

|                | 0.00000267*          | 0.00000225***        | 0.00000202***        |
|                | (0.00000139)         | (0.000000526)        | (0.000000497)        |

Observations 16422 16422 16422

Adjusted $R^2$ 0.621 0.621 0.621

Ecoregion: Dfa, temperature threshold: 19.

Table A.4: Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg ddayHigh</td>
<td>-0.00113***</td>
<td>-0.00111***</td>
<td>-0.00115***</td>
</tr>
<tr>
<td></td>
<td>(0.000115)</td>
<td>(0.000123)</td>
<td>(0.000110)</td>
</tr>
<tr>
<td>max ddayHigh</td>
<td>-0.00113</td>
<td>-0.00111</td>
<td>-0.00115</td>
</tr>
<tr>
<td></td>
<td>(0.000115)</td>
<td>(0.000123)</td>
<td>(0.000110)</td>
</tr>
<tr>
<td>p90 ddayHigh</td>
<td>-0.00113</td>
<td>-0.00111</td>
<td>-0.00115</td>
</tr>
<tr>
<td></td>
<td>(0.000115)</td>
<td>(0.000123)</td>
<td>(0.000110)</td>
</tr>
</tbody>
</table>

low ddays X exposure

|                | -0.000000281         | -0.000000253         | -0.000000222         |
|                | (0.000000233)        | (0.000000200)        | (0.000000193)        |
| high ddays     | -0.00158***          | -0.00164***          | -0.00160***          |
|                | (0.000125)           | (0.000132)           | (0.000130)           |
| high ddays     | -0.00158             | -0.00164             | -0.00160             |
|                | (0.000125)           | (0.000132)           | (0.000130)           |
| high ddays X exposure

|                | 0.000000434**        | 0.000000434**        | 0.000000420**        |
|                | (0.000000190)        | (0.000000178)        | (0.000000175)        |

Observations 51482 51482 51482

Adjusted $R^2$ 0.674 0.674 0.674

Ecoregion: Csb, temperature threshold: 12.
### Table A.5: Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>avg ddayHigh</td>
<td>max ddayHigh</td>
<td>p90 ddayHigh</td>
</tr>
<tr>
<td>low ddays</td>
<td>-0.00164***</td>
<td>-0.00159***</td>
<td>-0.00164***</td>
</tr>
<tr>
<td></td>
<td>(0.000335)</td>
<td>(0.000272)</td>
<td>(0.000308)</td>
</tr>
<tr>
<td>low ddays</td>
<td>X exposure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000000343</td>
<td>0.0000000289*</td>
<td>0.0000000329*</td>
</tr>
<tr>
<td></td>
<td>(0.000000213)</td>
<td>(0.000000159)</td>
<td>(0.000000190)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.000847*</td>
<td>-0.00114**</td>
<td>-0.000894**</td>
</tr>
<tr>
<td></td>
<td>(0.000490)</td>
<td>(0.000465)</td>
<td>(0.000421)</td>
</tr>
<tr>
<td>high ddays</td>
<td>X exposure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-8.76e-08</td>
<td>8.09e-08</td>
<td>-5.73e-08</td>
</tr>
<tr>
<td></td>
<td>(0.0000000299)</td>
<td>(0.0000000262)</td>
<td>(0.0000000249)</td>
</tr>
<tr>
<td>Observations</td>
<td>64574</td>
<td>64574</td>
<td>64574</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.548</td>
<td>0.548</td>
<td>0.548</td>
</tr>
</tbody>
</table>

Ecoregion: BSk, temperature threshold: 11.

### Table A.6: Influence of early exposure to degree-days above ecoregion-specific threshold (location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>avg ddayHigh</td>
<td>max ddayHigh</td>
<td>p90 ddayHigh</td>
</tr>
<tr>
<td>low ddays</td>
<td>-0.000458</td>
<td>-0.000202</td>
<td>-0.000186</td>
</tr>
<tr>
<td></td>
<td>(0.000779)</td>
<td>(0.000280)</td>
<td>(0.000360)</td>
</tr>
<tr>
<td>low ddays</td>
<td>X exposure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000000123</td>
<td>0.0000000561</td>
<td>0.0000000531</td>
</tr>
<tr>
<td></td>
<td>(0.000000191)</td>
<td>(0.0000000635)</td>
<td>(0.0000000824)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00403***</td>
<td>-0.00390***</td>
<td>-0.00414***</td>
</tr>
<tr>
<td></td>
<td>(0.000650)</td>
<td>(0.000464)</td>
<td>(0.000411)</td>
</tr>
<tr>
<td>high ddays</td>
<td>X exposure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-8.76e-08</td>
<td>8.09e-08</td>
<td>-5.73e-08</td>
</tr>
<tr>
<td></td>
<td>(0.0000000299)</td>
<td>(0.0000000262)</td>
<td>(0.0000000249)</td>
</tr>
<tr>
<td>Observations</td>
<td>183362</td>
<td>183362</td>
<td>183362</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.609</td>
<td>0.609</td>
<td>0.609</td>
</tr>
</tbody>
</table>

### A.3.9 Acclimatization: AMJJA temperature response

The following tables show estimates of the main coefficients from the acclimatization model applied to tree growth response to AMJJA temperature. The model is similar to the general year-round one described in the previous section.

Table A.7: Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1) avg pctile</th>
<th>(2) max pctile</th>
<th>(3) avg ddayHigh</th>
</tr>
</thead>
<tbody>
<tr>
<td>low ddays</td>
<td>-0.0009999**</td>
<td>-0.00118***</td>
<td>-0.00134**</td>
</tr>
<tr>
<td></td>
<td>(0.000159)</td>
<td>(0.000287)</td>
<td>(0.000364)</td>
</tr>
<tr>
<td>low ddays  X exposure</td>
<td>0.00000931</td>
<td>0.0000127</td>
<td>0.0000188</td>
</tr>
<tr>
<td></td>
<td>(0.0000100)</td>
<td>(0.0000111)</td>
<td>(0.0000143)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00684***</td>
<td>-0.00452</td>
<td>-0.00373</td>
</tr>
<tr>
<td></td>
<td>(0.00117)</td>
<td>(0.00300)</td>
<td>(0.00352)</td>
</tr>
<tr>
<td>high ddays X exposure</td>
<td>-0.0000934</td>
<td>-0.000144</td>
<td>-0.000179</td>
</tr>
<tr>
<td></td>
<td>(0.0000684)</td>
<td>(0.000119)</td>
<td>(0.000143)</td>
</tr>
</tbody>
</table>

Observations 4330 4330 4330

Adjusted $R^2$ 0.591 0.592 0.592


Table A.8: Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1) avg pctile</th>
<th>(2) max pctile</th>
<th>(3) avg ddayHigh</th>
</tr>
</thead>
<tbody>
<tr>
<td>low ddays</td>
<td>-0.00267***</td>
<td>-0.00269***</td>
<td>-0.00265***</td>
</tr>
<tr>
<td></td>
<td>(0.0000476)</td>
<td>(0.0000406)</td>
<td>(0.0000545)</td>
</tr>
<tr>
<td>low ddays  X exposure</td>
<td>-0.00000291</td>
<td>-0.00000858</td>
<td>-0.00000408</td>
</tr>
<tr>
<td></td>
<td>(0.000000768)</td>
<td>(0.00000281)</td>
<td>(0.00000547)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00398**</td>
<td>-0.00395***</td>
<td>-0.00390**</td>
</tr>
<tr>
<td></td>
<td>(0.00151)</td>
<td>(0.00145)</td>
<td>(0.00151)</td>
</tr>
<tr>
<td>high ddays X exposure</td>
<td>0.0000159</td>
<td>0.000000338</td>
<td>0.0000116</td>
</tr>
<tr>
<td></td>
<td>(0.0000300)</td>
<td>(0.0000150)</td>
<td>(0.0000219)</td>
</tr>
</tbody>
</table>

Observations 102207 102207 102207

Adjusted $R^2$ 0.649 0.649 0.649

Ecoregion: Dfb, temperature threshold: 27.
Table A.9: Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1) avg pctile</th>
<th>(2) max pctile</th>
<th>(3) avg ddayHigh</th>
</tr>
</thead>
<tbody>
<tr>
<td>low ddays</td>
<td>0.00278 (0.00171)</td>
<td>0.00162 (0.00104)</td>
<td>0.00127 (0.000961)</td>
</tr>
<tr>
<td>low ddays X exposure</td>
<td>-0.00000305 (0.00000185)</td>
<td>-0.00000157 (0.000000986)</td>
<td>-0.00000125 (0.000000964)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00193 (0.00185)</td>
<td>-0.00215*** (0.000693)</td>
<td>-0.00143** (0.000509)</td>
</tr>
<tr>
<td>high ddays X exposure</td>
<td>0.00000159 (0.00000197)</td>
<td>0.00000160** (0.000000612)</td>
<td>0.000000930* (0.000000456)</td>
</tr>
</tbody>
</table>

Observations: 16422 16422 16422
Adjusted $R^2$: 0.622 0.622 0.622

Ecoregion: Dfa, temperature threshold: 16.

Table A.10: Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1) avg pctile</th>
<th>(2) max pctile</th>
<th>(3) avg ddayHigh</th>
</tr>
</thead>
<tbody>
<tr>
<td>low ddays</td>
<td>-0.00218*** (0.000283)</td>
<td>-0.00218*** (0.000294)</td>
<td>-0.00215*** (0.000299)</td>
</tr>
<tr>
<td>low ddays X exposure</td>
<td>-0.00000133 (0.000000937)</td>
<td>-0.00000105 (0.000000808)</td>
<td>-0.00000127 (0.000000882)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00198*** (0.000357)</td>
<td>-0.00198*** (0.000360)</td>
<td>-0.00199*** (0.000370)</td>
</tr>
<tr>
<td>high ddays X exposure</td>
<td>0.000000937 (0.000000729)</td>
<td>0.000000780 (0.000000588)</td>
<td>0.000000875 (0.000000706)</td>
</tr>
</tbody>
</table>

Observations: 51482 51482 51482
Adjusted $R^2$: 0.664 0.664 0.664

Ecoregion: Csb, temperature threshold: 14.
Table A.11: Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>avg pctile</td>
<td>max pctile</td>
<td>avg ddayHigh</td>
</tr>
<tr>
<td>low ddays</td>
<td>-0.0129***</td>
<td>-0.0115***</td>
<td>-0.0132***</td>
</tr>
<tr>
<td></td>
<td>(0.00245)</td>
<td>(0.00268)</td>
<td>(0.00238)</td>
</tr>
<tr>
<td>low ddays</td>
<td>-0.00000130</td>
<td>-0.00000183</td>
<td>-0.00000110</td>
</tr>
<tr>
<td></td>
<td>(0.00000167)</td>
<td>(0.00000170)</td>
<td>(0.00000155)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00103</td>
<td>-0.00156**</td>
<td>-0.00105</td>
</tr>
<tr>
<td></td>
<td>(0.000798)</td>
<td>(0.000758)</td>
<td>(0.000670)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.000000107</td>
<td>0.000000136</td>
<td>-9.52e-08</td>
</tr>
<tr>
<td></td>
<td>(0.000000326)</td>
<td>(0.000000297)</td>
<td>(0.000000260)</td>
</tr>
<tr>
<td>Observations</td>
<td>64574</td>
<td>64574</td>
<td>64574</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.541</td>
<td>0.541</td>
<td>0.541</td>
</tr>
</tbody>
</table>

Ecoregion: BSk, temperature threshold: 4.

Table A.12: Influence of early exposure to degree-days above ecoregion-specific threshold (AMJJA, location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>avg pctile</td>
<td>max pctile</td>
<td>avg ddayHigh</td>
</tr>
<tr>
<td>low ddays</td>
<td>0.00341***</td>
<td>0.00348***</td>
<td>0.00352***</td>
</tr>
<tr>
<td></td>
<td>(0.00113)</td>
<td>(0.000504)</td>
<td>(0.000471)</td>
</tr>
<tr>
<td>low ddays</td>
<td>0.00000102</td>
<td>0.000000762</td>
<td>0.000000700</td>
</tr>
<tr>
<td></td>
<td>(0.00000316)</td>
<td>(0.00000135)</td>
<td>(0.00000128)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00617***</td>
<td>-0.00609***</td>
<td>-0.00640***</td>
</tr>
<tr>
<td></td>
<td>(0.000712)</td>
<td>(0.000494)</td>
<td>(0.000335)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.000000129</td>
<td>-0.000000137</td>
<td>-0.000000621</td>
</tr>
<tr>
<td></td>
<td>(0.000000164)</td>
<td>(0.000000870)</td>
<td>(0.000000645)</td>
</tr>
<tr>
<td>Observations</td>
<td>183362</td>
<td>183362</td>
<td>183362</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.610</td>
<td>0.610</td>
<td>0.610</td>
</tr>
</tbody>
</table>

Ecoregion: Cfa, temperature threshold: 25.
A.3.10 Acclimatization: AMJJA temperature response of temperature-sensitive trees

The following table shows estimates of the main coefficients from the acclimatization model applied to tree growth response to AMJJA temperature, and restricted to the sample of temperature-sensitive trees. The model is similar to the general year-round one described in two sections above.

Table A.13: Influence of early exposure to degree-days among T-sensitive trees (location FE in the interaction)

<table>
<thead>
<tr>
<th></th>
<th>(1) avg ddayHigh</th>
<th>(2) max ddayHigh</th>
<th>(3) p90 ddayHigh</th>
</tr>
</thead>
<tbody>
<tr>
<td>low ddays</td>
<td>0.00146***</td>
<td>0.00149***</td>
<td>0.00149***</td>
</tr>
<tr>
<td></td>
<td>(0.000187)</td>
<td>(0.000181)</td>
<td>(0.000180)</td>
</tr>
<tr>
<td>low ddays X exposure</td>
<td>0.0000124*</td>
<td>0.00000257</td>
<td>0.00000335*</td>
</tr>
<tr>
<td></td>
<td>(0.00000661)</td>
<td>(0.00000202)</td>
<td>(0.00000186)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00761***</td>
<td>-0.00772**</td>
<td>-0.00772**</td>
</tr>
<tr>
<td></td>
<td>(0.00294)</td>
<td>(0.00294)</td>
<td>(0.00295)</td>
</tr>
<tr>
<td>high ddays X exposure</td>
<td>-0.0000841***</td>
<td>-0.0000270***</td>
<td>-0.0000369***</td>
</tr>
<tr>
<td></td>
<td>(0.0000197)</td>
<td>(0.00000856)</td>
<td>(0.00000765)</td>
</tr>
</tbody>
</table>

Observations 95159 95159 95159
Adjusted $R^2$ 0.684 0.684 0.684

Temperature threshold: 32.
A.3.11 Acclimatization: decadal model

The following tables show estimates of the main coefficients from a decadal acclimatization model. The model is similar to the main acclimatization model, except that the impact of low degree-days and high degree-days is interacted with age decade dummies, and average exposure to supraoptimal temperatures during the previous decade (denoted “exposure” in the table below). As usual, the AMJJA model includes additional controls for winter temperature. Each column corresponds to a different ecoregion, with the temperature breakpoint used to split degree-days between low and high indicated in parenthesis.

Table A.14: Influence of previous decade exposure to degree-days above ecoregion-specific threshold

<table>
<thead>
<tr>
<th></th>
<th>(1) Dfc (21°C)</th>
<th>(2) Dfb (25°C)</th>
<th>(3) Dfa (19°C)</th>
<th>(4) Csb (12°C)</th>
<th>(5) BSk (11°C)</th>
<th>(6) Cfa (26°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>low ddays</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000256</td>
<td>0.000288**</td>
<td>0.000223</td>
<td>0.000203*</td>
<td>0.000373</td>
<td>-0.0000252</td>
</tr>
<tr>
<td></td>
<td>(0.000172)</td>
<td>(0.000116)</td>
<td>(0.000127)</td>
<td>(0.000101)</td>
<td>(0.000259)</td>
<td>(0.0000536)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00507***</td>
<td>-0.00312</td>
<td>-0.00118**</td>
<td>-0.000604**</td>
<td>-0.00167***</td>
<td>-0.000769***</td>
</tr>
<tr>
<td></td>
<td>(0.00123)</td>
<td>(0.00190)</td>
<td>(0.000440)</td>
<td>(0.000246)</td>
<td>(0.000585)</td>
<td>(0.000297)</td>
</tr>
<tr>
<td>low ddays X exposure</td>
<td>-0.00000161</td>
<td>0.000000571</td>
<td>-7.30e-08</td>
<td>0.0000001040</td>
<td>-0.000000183</td>
<td>0.000000499***</td>
</tr>
<tr>
<td></td>
<td>(0.0000162)</td>
<td>(0.00000548)</td>
<td>(0.00000119)</td>
<td>(9.19e-08)</td>
<td>(0.00000183)</td>
<td>(8.00e-08)</td>
</tr>
<tr>
<td>high ddays X exposure</td>
<td>-0.00000529</td>
<td>-0.0000182</td>
<td>0.00000142**</td>
<td>0.000000125</td>
<td>0.000000223</td>
<td>-0.000000301***</td>
</tr>
<tr>
<td></td>
<td>(0.00000612)</td>
<td>(0.0000193)</td>
<td>(0.00000621)</td>
<td>(0.00000176)</td>
<td>(0.000000593)</td>
<td>(0.000000847)</td>
</tr>
<tr>
<td>Observations</td>
<td>4398</td>
<td>103804</td>
<td>16686</td>
<td>52290</td>
<td>65572</td>
<td>186816</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.617</td>
<td>0.644</td>
<td>0.615</td>
<td>0.668</td>
<td>0.545</td>
<td>0.603</td>
</tr>
</tbody>
</table>

Table A.15: Influence of previous decade exposure to degree-days above ecoregion-specific threshold (AMJJA)

<table>
<thead>
<tr>
<th></th>
<th>(1) Dfc (21°C)</th>
<th>(2) Dfb (27°C)</th>
<th>(3) Dfa (16°C)</th>
<th>(4) Csb (14°C)</th>
<th>(5) BSk (14°C)</th>
<th>(6) Cfa (26°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>low ddays</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.000467</td>
<td>0.000362**</td>
<td>0.000612*</td>
<td>0.000398*</td>
<td>0.0104**</td>
<td>-0.000215</td>
</tr>
<tr>
<td></td>
<td>(0.000313)</td>
<td>(0.000172)</td>
<td>(0.000332)</td>
<td>(0.000213)</td>
<td>(0.000467)</td>
<td>(0.000166)</td>
</tr>
<tr>
<td>high ddays</td>
<td>-0.00401*</td>
<td>-0.00778***</td>
<td>-0.00185***</td>
<td>-0.000995**</td>
<td>-0.00216</td>
<td>0.000236</td>
</tr>
<tr>
<td></td>
<td>(0.00173)</td>
<td>(0.00248)</td>
<td>(0.000604)</td>
<td>(0.000411)</td>
<td>(0.00155)</td>
<td>(0.000400)</td>
</tr>
<tr>
<td>low ddays X exposure</td>
<td>-0.00000153</td>
<td>0.000000276</td>
<td>-0.000000241</td>
<td>0.000000471</td>
<td>-0.00000212</td>
<td>0.000000125***</td>
</tr>
<tr>
<td></td>
<td>(0.00000264)</td>
<td>(0.00000142)</td>
<td>(0.000000225)</td>
<td>(0.000000279)</td>
<td>(0.00000243)</td>
<td>(0.000000168)</td>
</tr>
<tr>
<td>high ddays X exposure</td>
<td>-0.0000241***</td>
<td>0.0000240</td>
<td>0.00000140**</td>
<td>0.000000330</td>
<td>0.000000140</td>
<td>-0.00000561***</td>
</tr>
<tr>
<td></td>
<td>(0.00000632)</td>
<td>(0.00000562)</td>
<td>(0.000000625)</td>
<td>(0.000000497)</td>
<td>(0.000000811)</td>
<td>(0.000000113)</td>
</tr>
<tr>
<td>Observations</td>
<td>4398</td>
<td>103804</td>
<td>16686</td>
<td>52290</td>
<td>65572</td>
<td>186816</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.618</td>
<td>0.644</td>
<td>0.617</td>
<td>0.660</td>
<td>0.541</td>
<td>0.605</td>
</tr>
</tbody>
</table>
Figure A.19: Adaptation map: impact of degree-days above 23°C in humid ecoregions (all grid cells). Impact on annual tree growth of an additional degree-day above 23°C in humid ecoregions. The model is run separately on each 1° grid cell.
A.3.13 Adaptation: various subsamples

Figure A.20: Adaptation maps: frequency and impact of degree-days above 26°C in region Cfa (selected grid cells). [Left] Average number of annual degree-days above 26°C, by 1° grid cell. [Right] Impact on annual tree growth of an additional degree-day above 26°C, in region Cfa. The model is run separately on each 1° grid cell. Only negative and statistically significant coefficients are displayed, for graphical clarity.

Figure A.21: Adaptation scatterplot: impact of degree-days above 26°C as a function of their frequency, in region Cfa. The units of analysis are 1° grid cells.
Figure A.22: Adaptation maps: frequency and impact of degree-days above 17°C among Douglas Fir trees (selected grid cells). [Left] Average number of annual degree-days above 17°C, by 1° grid cell. [Right] Impact on annual tree growth of an additional degree-day above 17°C, among Douglas Fir trees. The model is run separately on each 1° grid cell. Only negative and statistically significant coefficients are displayed, for graphical clarity.

Figure A.23: Adaptation scatterplot: impact of degree-days above 17°C as a function of their frequency, among Douglas Fir trees. The units of analysis are 1° grid cells.
Figure A.24: Adaptation maps: frequency and impact of degree-days above 14°C among Ponderosa Pine trees (selected grid cells). [Left] Average number of annual degree-days above 14°C, by 1° grid cell. [Right] Impact on annual tree growth of an additional degree-day above 14°C, among Ponderosa Pine trees. The model is run separately on each 1° grid cell. Only negative and statistically significant coefficients are displayed, for graphical clarity.

Figure A.25: Adaptation scatterplot: impact of degree-days above 14°C as a function of their frequency, among Ponderosa Pine trees. The units of analysis are 1° grid cells.
Figure A.26: Adaptation maps: frequency and impact of degree-days above 23°C among White Oak trees (selected grid cells). [Left] Average number of annual degree-days above 23°C, by 1° grid cell. [Right] Impact on annual tree growth of an additional degree-day above 23°C, among White Oak trees. The model is run separately on each 1° grid cell. Only negative and statistically significant coefficients are displayed, for graphical clarity.

Figure A.27: Adaptation scatterplot: impact of degree-days above 23°C as a function of their frequency, among White Oak trees. The units of analysis are 1° grid cells.
Chapter 2

Monitoring shifts in flowering phenology using satellite imagery and deep learning

with Mathias Lécuyer
Abstract

Shifts in the timing of plant flowering are a key signal of an ecosystem’s response to environmental change. Consequences are far-reaching: the timing of plant flowering affects the synchronization between plants and pollinators; it influences the exposure of crops to weather extremes during flowering, the most sensitive stage of development, with implications for later yield. A consistent and large-scale measure tracking the timing of flowering across years would be of great use for assessing risks of plant-pollinator desynchronization or risks of subpar crop yields. Current efforts, ranging from local to regional scales, do not quite achieve this goal: ground observations, while expanding, remain limited in their spatial and temporal coverage; process-based or statistical models are not flexible enough to capture local acclimation or adaptation to environmental factors; and remotely sensed estimates of plant development stages are often disconnected from the ground. Hence, crop studies still rely on a time-invariant harvesting calendar to infer the flowering period for crops. We propose a method that combines spectral reflectance data from satellite imagery, with a deep learning model, to create a large-scale, high-resolution proxy for the timing of flowering. This method has two key components: it overcomes ground truth sparsity by first learning to predict a temperature proxy for flowering, and it learns from the temporal dynamics of spectral reflectance data, rather than from spatial patterns. We illustrate this approach with two use cases. First, over cropland in Illinois and Iowa, we construct a proxy measure of the evolution of crop flowering—silking for corn, flowering for soybean—across the growing season. This measure improves on a null average model and estimates from the literature, and captures some of the interannual variability in the timing of crop flowering. Second, over the Eastern US, we construct a proxy measure of the onset of spring flowering, which reflects the phenology of early-flowering, temperature-sensitive species. This measure improves over the current temperature-based first bloom index from the National Phenology Network. Again, it captures some of the interannual variability in the onset of spring flowering. Our proposed approach, despite using coarse satellite inputs and restricted ground truth data on flowering, holds great promise as more of those data sources become available.
2.1 Introduction

The onset of plant flowering captures the start of a critical period. Plants and pollinators engage in an essential mutualistic interaction, crops become particularly sensitive to environmental factors, and pollen allergies soar. The timing of this phenological event varies from year to year, especially in the context of climate change (Anderson et al., 2012). Capturing those interannual variations in the timing of flowering across large scales has numerous applications. First, it could be instrumental for understanding the evolution of synchronization between plants and pollinators. There is evidence that pollinators and plants respond to different environmental cues, and might become de-synchronized under climate change (Solga et al., 2014), with implications for both populations. Existing evidence is mixed, and constrained by lack of co-occurring data on plant flowering and pollinator flight. A consistent large-scale proxy for the local onset of flowering would shed light on plant-side shifts in phenology, and allow further matching to existing pollinator observations. Second, in agricultural lands, flowering is a key stage: it captures a period of time when crops are most sensitive to environmental factors (Siebers et al., 2017; Ortiz-Bobea and Just, 2013; Barnabás et al., 2008). Large-scale studies that estimate the impact of climate on yields use a time-invariant crop calendar (Sacks et al., 2010). This calendar does not take into account the fact that the timing of the sensitive flowering period is itself shifting under climate change, despite intense management (Anderson et al., 2017). This could lead to underestimation of climate change impacts on yields (Lobell et al., 2012; Duncan et al., 2015b). Third, in urban settings, capturing interannual variations in the timing of spring flowering would help shed light on the connection between allergies, urban vegetation, and climate (Sapkota et al., 2019).

Current efforts to monitor the onset of spring flowering at regional to continental scales rest on numerous but sparse ground observations (Ovaskainen et al., 2020; Rosemartin et al., 2015; Wolkovich et al., 2012), on a temperature-based proxy (Schwartz et al., 2006), or on remotely sensed vegetation metrics, often disconnected from the ground (Cleland et al., 2007). In crop studies, a time-invariant harvesting calendar is most widely used to infer the flowering stage for crops at large scales (Sacks et al., 2010). Several methods have also been developed to infer the timing of crop development stages from remotely sensed vegetation indices (see Duncan et al., 2015a, for a review). These methods usually

---

1In general, animal phenology is driven more by temperature and less by photoperiod or water, relative to plant phenology (Chuine and Régnière, 2017). For example, pollinators time their flying depending on the temperature of the air, to maximize egg and larval survival, while plant flowering responds more strongly to snow melt and photoperiodicity.

2For example, a warmer winter could trigger flowering before the last frost occurs; a late start of season could induce a shorter growing season and drought (Meroni et al., 2014, Horn of Africa); earlier planting could decrease exposure to deleterious high temperatures (Lobell et al., 2013, India).
start with pre-processing of satellite images, to minimize contamination by clouds and other sources of noise; choice of a vegetation index; application of a smoothing curve to the time series of the vegetation index; and finally retrieval of crop phenological transition dates using algorithms based on inflection points or thresholds, sometimes in conjunction with crop growth models and weather data. However, there remains a gap between those remotely sensed estimates and ground-level crop physiological development, due to lack of model validation. Field-level observations of crop phenology are limited in time and space, particularly for the flowering stage, which makes remotely sensed phenology sequences an approximation at best.

In this chapter, we combine MODIS satellite imagery with a deep learning model to construct a large-scale proxy measure of the timing of flowering. We seek to construct a measure that satisfies the following criteria: (1) availability at a relatively fine spatial resolution to capture local variations, (2) ability to capture interannual variations in the timing of flowering, and (3) ability to extrapolate across time and space, to map flowering across many years and large geographical areas. Our measure has the advantage to not require plant-specific knowledge on physiological growth. This is because deep learning models learn model features rather than estimate pre-specified ones. We specifically choose a temporal neural network architecture, so that the model learns temporal patterns in spectral reflectance data that are predictive of the onset of flowering.

Flowering occurs at the level of an individual plant, which prompts the question of what it means to measure flowering at a coarser scale. We address this question in two different ways. First, we look at cropland: flowering is relatively synchronized within fields, so it can be extrapolated to a field-level phenomenon without excessive distortion. Over cropland in Illinois and Iowa, we construct a proxy measure of the evolution of crop flowering—silking for corn, flowering for soybean—across the growing season. Second, we track a flowering event that can be defined at the landscape level: flowering is highly heterogenous across individual plants, and stretches throughout the year, so we focus on the first significant emergence of flowers prior to leaf-out, which marks the arrival of spring in temperate regions. Over the Eastern US, we construct a proxy measure of the onset of spring flowering, which reflects the phenology of early-flowering temperature-sensitive species.

Both our models show that combining satellite imagery with deep learning holds great promise for monitoring different metrics of flowering phenology. Our crop model predicts the evolution of crop flowering quite closely, capturing important interannual variations at a sub-state level (called agricultural statistics districts) and at the local level of MODIS pixels. The timing of 50% corn silking
and 50% soybean flowering is predicted at the agricultural district level with an out-of-sample root mean squared error (RMSE) of 5.2 and 5.9 days respectively, which is strictly less than the null model RMSE of 8.1 and 7.8 days, and captures some interannual variation at the district level (the standard deviation of the timing of this event is 7.7 days on average across this test set). The exact error is sample-dependent: our model error is much lower on validation sets that have lower interannual variability. This is an improvement over existing crop flowering models, but our model does not capture the full extent of interannual variations in the timing of crop flowering, especially in Northern and Western Iowa. Selecting subsets of satellite bands reveals that the model learns best when all bands are included. Model learning is particularly dependent on the inclusion of the red band, either alongside the rest of the visible range, or alongside bands in the near infrared. When bands in the visible range are included, adding bands in the near infrared, bands in the short-wave infrared, or non-spectral bands, impacts model performance marginally.

Our model of spring onset predicts the timing of first flowering for early-flowering temperature-sensitive species. It improves over the current temperature-based first bloom index, produced by the National Phenology Network, even when we compare our out-of-sample error to their in-sample error over our core study region. Performance is heterogeneous across space: at most locations, model error is low enough (under 8 days) to capture interannual variation in the timing of spring onset, whereas at a few locations, the error is too high. Most importantly, our model extrapolates much better over space from our core study region to the broader Eastern US, than the temperature-based first bloom index. This suggests that remotely sensed features identified using deep learning, can alleviate some limitations of weather-based models, that usually do not take local acclimation and adaptation into account. Such features can be used to construct proxies of the onset of flowering across large areas and across years, which could greatly complement ground-level, statistical-based, and process-based approaches.

2.2 Related Work

**Crop flowering.** The literature is rich in remotely-sensed crop phenology estimates, but rare in ground-validated ones. The closest study to ours, Diao, 2020, uses a 16-day MODIS product to predict corn and soybean growth stages over Illinois in 2002-2017, using several curve-fitting-based phenological models and phenophase estimation methods. For each model, they predict the date of peak silking for corn and peak blooming for soybean at the pixel level, and aggregate those predictions
by agricultural district, to create district-level curves of the cumulative corn and soybean flowering progress. The RMSE of their best performing model is about 10 days for corn silking (the worst estimated stage for corn) and 5 days for soybean blooming (the other models that they test all have an error higher than 20 days for this stage). For reference, variability in flowering across years is 6 days on average in their sample for both corn and soybean, based on the standard deviation of the timing of median flowering. All other studies assess their predictions against coarser state-level crop progress data, or sparse experimental plots. For example, Wang et al., 2020a, use MODIS satellite data and a rule-based rice phenology detection algorithm to determine the timing of rice phenology stages. Compared to 16 years of California-level data on rice growth progress, their predictions of heading\(^3\) have a RMSE of about 6 days. This stage is the least precisely estimated, compared to planting and harvest. Sakamoto et al., 2005, also predict the timing of rice heading, with a RMSE of about 10 days relative to statistical records from Japanese cropping zones. Bandaru et al., 2020, predict corn silking and soybean blooming in Nebraska. Their state-level RMSE is 1.1 days for corn and 6.7 days for soybean. However, their state-level predictions are likely overfitted: they use state-level data on crop phenology to fit parameters from a photothermal model, which is used in combination with normalized difference vegetation index (NDVI) curves to infer the timing of each stage, so they use in input the outcome that they are trying to predict; there is no split between a training and a testing set; and the R squared of their model is 0.99. They also evaluate model performance against local level crop flowering, using data from a couple of experimental fields. The field-level error is 6.5 days for corn silking and 4.8 days for soybean flowering, although it cannot be directly compared with our district-level errors. Across these studies, it is important to keep in mind that the magnitude of prediction errors has to be compared with the underlying variability in the data sample, which predictions are trying to capture. A smaller error does not necessarily denote a better performing model, if it is estimated on a less variable sample.

**Onset of spring flowering.** Flowering phenology has been studied at a range of geographic and temporal scales. In particular, large-scale and long-term patterns of the onset of spring flowering are informed by a variety of tools. First, in situ observations are key and many long-term data collection efforts expand over large geographic regions. They include but are not limited to: the National Ecological Observatory Network (Elmendorf et al., 2016), the USA National Phenology Network (Rosemartin et al., 2015; Denny et al., 2014), the Network of Ecological and Climatological Timings Across Regions

\(^3\)They don’t predict flowering. Heading is the closest stage, it occurs right before the flowering stage.
database (Wolkovich et al., 2012), which encompasses several Long Term Ecological Research Network sites, the Pan European Phenological database (Templ et al., 2018), the recent Chronicles of nature calendar in and around Russia (Ovaskainen et al., 2020), and long-term historical records from Asia and Europe (reviewed in Cleland et al., 2007). There are also ongoing efforts in digitization of herbarium data (Willis et al., 2017; Yost et al., 2018; Love et al., 2019) and citizen science contributions like iNaturalist. While immensely useful, these efforts are species and site specific, or lack temporal coverage in the case of herbarium and citizen science data. We test whether remote sensing can exploit and complement such ground data observations to provide more complete coverage at a large scale.

Second, process-based phenology models (reviewed in Chuine and Régnière, 2017), inform the response of physiological processes to specific factors like temperature or photoperiod. They provide invaluable mechanistic understanding, but have an important drawback. Parameters usually do not vary by geography to allow local adaptation to climate, and models do not allow acclimation to temporally variable environmental factors. Our model relies on extracting visual signals rather than on parametrizing underlying physiological processes, which we hope can have more external validity.

Third, remote sensing has been used to capture spring phenology, using a variety of spectral indices and spring onset metrics. White et al. (2009), review different NDVI-based “start-of-spring” models. Their interest is primarily in comparing predictions across models, and estimating trends in spring onset. Hence they use a relatively coarse 15-day NDVI composite at an 8 km resolution, which performs poorly for predicting ground level observations of first flowering across all models. In general, remotely sensed predictions of spring flowering onset lack ground-level validation and target leaf stages more frequently than flowering stages, as is the case for remotely sensed predictions of crop phenology. As an example of connexion across scales, Liang et al. (2011), show that MODIS-derived EVI performs much better than NDVI for monitoring locally measured budburst at two locations over two years.

There exists one large-scale dataset documenting the onset of spring flowering across the contiguous US. This “first bloom index” is constructed using temperature data at a 4 km resolution, and latitude-derived day length (Schwartz, 1997; Schwartz et al., 2006, 2013; Ault et al., 2015b,a). It is parametrized using long-term observations of lilac and honeysuckle flowering (Rosemartin et al., 2015). As with process models, this temperature model does not allow different temperature sensitivities, other than by latitude. We take this first bloom index as a reference, and test whether a remote sensing approach can improve on this model’s performance and extrapolation power.
2.3 Core Approach

2.3.1 Idea: exploit convolutional neural networks and transfer learning

Flowering is, among other things, a visual event. Our intuition is that there is a sharp temporal transition in what a landscape looks like around the onset of flowering. This could be captured by satellite imagery, which consists of time series of images across locations. We expect there to be characteristic transitions in satellite-based spectral reflectance that could capture the onset of flowering. It is likely difficult to characterize a priori the exact nature of these transitions. It is easy to picture what a superbloom looks like over the Carrizo Plain Monument in California (Figure 2.1), but many transitions are likely more subtle, involve some of the non-visible spectral bands, and are overall varied. This is why we turn to deep learning.

A deep neural network is essentially a cascading set of non-linear functions. We design our own network to have convolutions in the time dimension, so that the model learns temporal patterns across the different spectral bands, which matter for predicting the timing of flowering. As an example, a temporal pattern could be an increase, break-point, or change in the curvature of one or a combination of spectral band time series. There is no need to specify these patterns beforehand; the model learns them by itself. However, it is a supervised model, which means that it is trained using ground-truth points where the value of the “label” (the flowering event we want to predict) is known.

Complex supervised models come at a cost: estimating the many parameters of the model requires a great amount of ground truth data. There exists field observations of flowering, but they are quite sparse and difficult to harmonize across different sources. To overcome this problem, we use a classic technique called transfer learning (Pan and Yang, 2009). We first train the model to predict an
intermediate outcome, which is related to flowering, but for which there exists a lot more ground truth data. This step helps move the model parameters from a random initialization value into a relevant range, as the model starts learning temporal features, which are relevant for predicting a rough proxy for the onset of flowering\textsuperscript{4}. We then fine-tune the model by adjusting some of the parameter values, using the more sparse ground observations on flowering. In practice, we implement transfer learning using temperature data. It is a good candidate for an intermediate outcome because it is widely available and an important driver of flowering. For example, the timing of corn phenostages is closely related to air temperature (Abendroth et al., 2011), while soybean phenology is particularly driven by air temperature and photoperiod (Setiyono et al., 2007). We do not use photoperiod for transfer learning because it is fixed from one year to the next so it will not help the model learn features that can explain interannual shifts in flowering drivers\textsuperscript{5}.

2.3.2 Model architecture

Figure 2.2 describes the architecture of the model. The task assigned is the following: for a given pixel and a given year $t$, the model takes a sequence of satellite observations of the pixel, spanning $T$ dates in year $t$; the satellite observations contain information on $N$ spectral bands; the model uses this input to predict a flowering event at that pixel in year $t$. The event can be a single point in time, like the date of first flowering, or a times series, like flowering progress over time in year $t$, at each of the $T$ dates. The transfer learning stage simply corresponds to a temporary change in the predictive goal: at first, the model predicts a temperature-based event or a temperature sequence in year $t$, before moving on to predicting a flowering event or a flowering sequence.

The model contains a sequence of linearly connected hidden layers of neurons, interspersed with non-linear transformations: “ReLU”, or Rectified Linear Unit, activates a subset of neurons of a layer by running the function $\max(0,x)$ on each neuron’s value; “Dropout” also activates a subset of neurons of a layer by dropping entire features with some probability value during training. These non-linear transformations prevent the entire model from collapsing into a linear function.

The key feature of this architecture is the temporal convolution outlined in the bottom right inset. The vertical bars on the left represent the input, for example the sequence of values of red, green,

\textsuperscript{4}Parameter values can also be initialized by taking the parameter values from a similar pre-trained model from the literature. In our case, there doesn’t exist a temporal neural network model trained on multispectral data for a regression task, hence the random initialization.

\textsuperscript{5}In addition, photoperiod is correlated with solar zenith angle, which is one of the satellite inputs to the model, so the model would likely focus on learning this correlation, which will not help towards predicting flowering.
Figure 2.2: **Core model architecture.** Temporal convolutional network, adapted from Bai et al. (2018). [Top] Overall architecture. [Bottom left] A dilated block. A 1-convolution is included when input and output have different dimensions. The model includes 3 such dilated blocks. [Bottom right] Illustration of a dilated convolution with dilation factors $d = 1, 2, 4$ and filter size $k = 3$.

and blue reflectance across $T$ dates in year $t$. The red rectangle is a sliding window that takes $k = 3$ observations and links them to one point (neuron) in the next layer, then the window slides down by one square and links its 3 inputs to the next point in the next layer, etc. This essentially defines relationships between the output and input. Here we use dilated convolutions: in the first dilated block, the window links 3 consecutive observations (it skips $d - 1 = 1 - 1 = 0$ time points between each observation); in the second dilated block, the window links 3 observations after skipping $d - 1 = 2 - 1 = 1$ time point between each observation; and in the third dilated block, the window links 3 observations after skipping $d - 1 = 4 - 1 = 3$ time points between each observation. This gives longer memory to the model: it learns short-term temporal features, for example features based on 3 consecutive observations of the satellite product, and temporal features that are more spread out over a year.

A few parameters and architecture pieces depend on the model’s use case. The number of bands $N$ depends on the dimensions of the satellite imagery used as input ($N=13$ in both use cases presented in this chapter). The number of dates $T$ in the input sequence depends on the temporal frequency of the input imagery and on the time period used for prediction in each use case. The dimension of intermediate layers, including parameter $M$, is flexible and thus also varies with the use case. The yellow block depends on the exact predictive task. It connects the output of the last dilated block
(an M x T matrix, transformed from the initial N x T input through a series of dilated blocks), to the output of the model, which is a single number or a sequence of numbers, depending the flowering event targeted.

2.3.3 Model evaluation

To evaluate model performance, the data are split between a training set, a validation set, and a test set. We randomly assign entire years to each set, to assess whether the model can extrapolate to years “unseen” in the data, without relying on spatial correlation in phenology. This is because we are primarily interested in capturing interannual variations in the timing of flowering at any given location. During model development, model performance is assessed on the validation set. Once the model structure is set, the model is trained on the combined training and validation sets, and evaluated on the test set. We report this final performance evaluation, using the root mean squared error (RMSE) metric.

2.3.4 Input data: satellite imagery

The input data to our models is the MODIS Terra Surface Reflectance 8-Day Global 500m product (DAAC, 2021), which we download using Google Earth Engine’s Python API. It includes 7 reflectance bands, spanning the visible to the short-wave infrared, 2 quality control masks, solar zenith angle, view zenith angle, relative azimuth angle, and day of year\textsuperscript{6}, which makes 13 bands in total. Despite its relatively coarse spatial resolution, this product has several advantages over other satellite data sources: it is a regular 8-day product that filters images for quality (it selects the least cloudy, aerosol-loaded, and zenith-angled images within each 8-day period); it covers a large spectral range, extending from the visible to the short-wave infrared; it is publicly available; and most importantly, it covers the period 2001-2018, which is long enough to capture interannual variations in flowering.

We do not further process the satellite product for clouds. It has been shown that convolutional neural networks can learn a cloud-filtering mechanism by reducing the weight on cloudy observations, when input data contain both cloudy and non cloudy observations (Rußwurm and Körner, 2018). More generally, the advantage of deep learning models is that they can learn to recognize various sources of noise that impact spectral reflectance data, which circumvents the need for data preprocessing.

\textsuperscript{6}This day of year variable is included as a predictor in the model, as a way to precisely date the MODIS image time series used as input for each pixel and year (instead of relying on the coarser 8-day resolution). From this, the model predicts a time series of flowering progress throughout the year, or the date of a single flowering event in the year. In any case, we are not using the “date of flowering” as a predictor in our model.
We focus on extracting temporal features from this satellite imagery. White et al. (2009), assessed how well different NDVI-based algorithms matched ground level observations related to the start of spring, including first flowering and first leafing. They found that the two algorithms that best match first flowering observations, focus on capturing a rate of change in the NDVI time series, as opposed to a fixed threshold-exceedance, which suggests focusing on temporal changes in reflectance data. More recently, Garnot et al. (2019), assessed the relative importance of spatial and temporal structure in deep learning models, for crop type classification based on multi-spectral time series. They find that the best performing model is one that allocates 90% of the model parameters to the temporal structure of the data. This is a classification problem, unlike ours, but differences in temporal signature that distinguish different crops is likely to come from differences in phenology, which is exactly what we are trying to capture.

2.4 Use Case 1: Predicting Crop Flowering Progress

2.4.1 Model

In the first use case, we predict corn and soybean flowering progress over the growing season, across the states of Illinois and Iowa. Figure 2.3 shows how the architecture of the core model is adapted to the crop flowering task. We restrict the input sequence to the months of April to October, which enclose the entire crop growth season with a buffer. Combined with the 8-day frequency of the MODIS series, this implies an input sequence of $T = 27$ time points. The output of the model is a sequence of the same length, where the percentage of flowering is predicted at each date. The output of the last dilated block is connected to the final output by a 1-convolution, which maps linearly a $(32 \times 27)$ matrix into a $(1 \times 27)$ vector. For the transfer learning stage, we predict the sequence of weekly average temperature, at the $T = 27$ time points.

To stabilize model predictions, we create an ensemble model: we train three independent models that have the same architecture, and average their predictions.
2.4.2 Data

2.4.2.1 Ground level crop phenology

Crop phenology data for Illinois and Iowa in 2001-2017 come from the USDA NASS, which publishes weekly crop progress reports for corn and soybean, aggregated at the agricultural statistical unit level\(^7\). We use corn silking progress and soybean blooming progress to capture the evolution of flowering (see appendix B.1.1 for more details on these metrics). We convert the raw weekly frequency into a daily one by linear interpolation, since cumulative flowering is a monotonous function of time. It yields relatively smooth and often logistic-like curves, a shape that is always apparent from the raw weekly data. We define the 50%-flowering event as the time when cumulative flowering progress has reached 50% in a given district and year. We use this 50%-flowering event to expose more clearly interannual variation in the timing of flowering, and differences between predicted and observed flowering progress.

In order to match district-by-crop flowering data with higher resolution satellite data, we use agricultural statistical district boundaries from a 5 m resolution USDA NASS shapefile, and annual 30 m corn-soy classification from Wang et al. (2020b). This classification, called the Corn-Soy Data Layer, is a retrained version of the Cropland data Layer (USDA-NASS, 2021). Using Landsat imagery, the updated classification extends the Cropland Data Layer back in time, to homogenize the spatial coverage and the underlying classification methodology. This new classification performs very well in Illinois and Iowa, and generates maps with less local noise than the original Cropland data Layer (see figure 1 in Wang et al., 2020b).

2.4.2.2 Weather data for transfer learning

For the transfer learning step, we use daily weather data from PRISM (Daly et al., 2008, 2015), which we download using Google Earth Engine’s Python API. We average weather variables over consecutive 8-day periods to match the temporal resolution of the MODIS product. For any given 8-day period, we choose average weather, rather than weather on the exact date of that period’s MODIS image, to capture weekly temperature accumulation rather than daily fluctuations. The model will learn temporal features in spectral reflectance, which help predict the evolution of temperature between April and October. We hope that these temporal features will capture something about the speed of seasonal evolution, which will be relevant for predicting the evolution of crop flowering. In addition to

\(^7\)Only state level aggregates are publicly available. We obtained district level data for Illinois and Iowa by contacting state and regional offices of USDA NASS.
weather, crop management exerts an important influence on the timing of crop growth development, which could hinder the model’s ability to predict crop flowering. However, management decisions are likely partially driven by interannual differences in temperature, and the fine-tuning stage of model training can help the model “learn” additional features reflecting this management influence (for example, spectral cues indicating temporal shifts in planting and emergence).

2.4.2.3 Crop study region and summary statistics

Figure 2.4 shows our study region for the crop flowering product on the left. It covers the states of Illinois and Iowa, which contain 18 agricultural statistical districts in total. The maps on the right show the data in more detail over one rectangular inset. From the map on the top left to the map on the top right, the cropland data layer is reduced to the resolution of MODIS, whose pixels are outlined in black. A MODIS pixel is categorized as cropland as soon as more than 50% of its area is covered by cropland; within cropland, a pixel is categorized as corn, soybean or other based on the dominant crop type represented within its area. The map on the bottom left shows data from MODIS at its original resolution, and as a true color composite. And the map on the bottom right shows mean temperature data from PRISM, upscaled to match the spatial resolution of MODIS, and averaged over the corresponding 8-day period. The average farm size in Illinois and Iowa in 2018 was 375 acres and 356 acres respectively, which represents about 6 MODIS pixels. Spatial resolution of the input spectral reflectance data is thus well adapted to monitoring crop phenology in this region.

Figure 2.5 gives an example of the raw input data, over a randomly selected corn field in Iowa. It shows the evolution of the normalized difference vegetation index (NDVI), a measure of “greenness”, over the year 2018, along with the evolution of daily mean temperature from PRISM. This figure highlights the common seasonality between the two signals. Plant phenology is tightly coupled with temperature, which supports using temperature as our transfer learning target.

Finally, figure 2.6 shows the extent of interannual fluctuations in the timing of corn and soybean flowering, which we hope to capture with our model8. It shows, for each agricultural statistical district, the variability in the date on which 50% flowering is reached in that district, across our study period (2001-2017). The width of the interquartile range varies between 4 days and about two weeks. To get a more precise estimate of interannual variability, we compute the standard deviation in the 50%-flowering date for each district and crop. This standard deviation ranges between 4.6 days and 9.7

8Appendix B.2.1 shows crop flowering progress curves in more detail.
Figure 2.4: Crop model study region and sample maps of cropland, spectral reflectance, and temperature data. The map on the left shows our study region, which covers the states of Illinois and Iowa. The boundaries of each agricultural statistical district are outlined in black. Colors represent land cover assignment in 2018, based on the Corn-Soy Data Layer: yellow is for corn, green for soybean, brown for other crop, and purple for non cropland. The four maps on the right show our data sources over a smaller region. Across all maps, the boundaries of MODIS pixels are outlined in black. The top left map shows land cover assignment in 2018, based on the Corn-Soy Data Layer, at the native 30 m resolution. The top right map shows the same land cover assignment, summarized at the resolution of MODIS pixels. The bottom left map shows imagery from MODIS for the 8-day period starting on June 2nd, 2018. It uses the true color band composition (Bands 1, 4, 3 — Red, Green, Blue). The bottom right map shows the average temperature for the 8-day period starting on June 2nd, 2018, based on PRISM data. The native resolution is 4 km, hence the homogenous patches. Colors range from cyan to red, for a corresponding temperature range from 22.7°C to 22.9°C, to highlight the coarser native resolution compared to MODIS.

days across district-crop pairs, and equals 6.3 days on average\(^9\). Very importantly, this interannual variability varies across sample periods. When the time period is restricted to years present in the validation set, the average standard deviation is equal to 4.3 days (4.8 days for corn, 3.9 days for soybean). When the time period is restricted to years present in the test set, the average standard deviation is equal to 7.7 days (7.8 days for corn, 7.7 days for soybean). Since model performance is assessed on the test set, the latter figures are the relevant reference numbers for interannual variability: a useful prediction model should have a RMSE strictly under 7.7 days, to capture an important share of interannual variability.

\(^9\)This means that, on average, a 6.3-day interval around the average 50%-flowering date is needed, to capture the 50%-flowering date in 60% of the years, for a given district and crop.
Figure 2.5: Example of annual evolution of NDVI and temperature, over a randomly selected crop field. This plot shows the evolution over the year 2018 of NDVI from MODIS at an 8-day resolution, and average daily temperature from PRISM, over one crop field in Iowa. The exact location of the crop field is (41°48'52.0"N 93°34'37.3"W).

Figure 2.6: Interannual variability in the timing of crop flowering. Variability in the date on which 50% of flowering is reached, by agricultural district and crop type (corn on the left, soybeans on the right). The data cover the period 2001-2017.
As an additional reference, we assess the performance of a null model, which uses the average timing of 50%-flowering for each district and crop type, to predict the timing of this event in every year. This null model, trained on the combined train and validation sets, has a RMSE of 7.9 days on the test set (8.1 days for corn, 7.8 days for soybean). Our model has to have a strictly lower RMSE, in order to be more informative than this null model.

2.4.3 Results

The input to the crop model is at the MODIS resolution, whereas ground observations used to validate the model are at the agricultural district level. The model thus first predicts crop flowering progress at each MODIS pixel, and then aggregates across pixels within a district to create crop flowering progress curves at the agricultural district level. District-level comparisons of observed and predicted progress drive model learning during the training stage. Using flowering progress curves, we compute the date at which 50% flowering is reached for each crop, district, and year.

Over all districts and test years, the RMSE of the 50% flowering prediction is 5.2 days for corn and 5.9 days for soybean, which is strictly lower than the close-to-8-day target set in the previous section. Figure 2.7 shows a scatterplot of predicted dates against observed dates of 50% flowering, by district and test year, and split by crop. Overall, predicted dates are less spread out over time than observed dates are: the model predicts late corn silking too early, and predicts early and late soybean flowering too late and early, respectively. Model performance differs by state: in Illinois (dots), predicted dates closely match observed dates (dots are close to the 45° line), whereas in Iowa about half of the predicted dates are far from the target (triangles are clustered within a narrow horizontal band). Different colors denote different years. There are no obvious patterns by year, aside from the fact that there is temporal correlation in the timing of flowering (see for example earlier flowering in 2007 and later flowering in 2008).

In appendix B.2.2, figures B.3 and B.4 show the same information, but displayed so that district-level interannual variations are apparent: each panel grid cell represents a district, so comparisons across test years within grid cells show whether the model is capturing interannual variations. Model performance is quite variable across districts. In many districts, like district 1730, the model prediction closely tracks interannual variations in the event date for both crops. In other districts, like district

---

10. In comparison, when the train set is used for training, the RMSE on the validation set is 4.4 days (4.6 for corn, 4.1 for soybean), which highlights the importance of drawing the right comparison.

11. In reality, for computational reasons, the model always aggregates over a small random subset of pixels (called a “batch”) at each training step. It only aggregates over the complete set of pixels at test time.
Given our focus on capturing interannual variation in the timing of flowering, we dig into heterogeneity in performance at the district level. We find that there are five districts for which our model does not beat the null model: they correspond to all the districts on the Northern and Western borders of Iowa (districts 1910, 1920, 1930, 1940, and 1970). The clear geographic clustering of those districts suggests that an environmental factor, such as greater cloud occurrence, could be hampering model learning. In the remaining districts, the RMSE of the 50% flowering prediction is 4.1 days for corn and 4.7 days for soybean (against a null model performance of 9.5 days and 8.9 days, respectively).

The model predicts cumulative flowering progress across the growing season, not just the timing of 50% flowering, and these progress curves are included in appendix B.2.3. They show that predicted flowering progress (dashed curves) tend to have a steeper slope than observed flowering progress (solid curves). This could be because the model predicts more synchronous progress across pixels within a district, relative to true progress synchronization. This is consistent with the previous observation that the model does not capture well extreme shifts in timing, so likely does not capture well the phenology of pixels that experience the earliest or the latest flowering. These flowering progress curves
confirm that the model seems to be performing better over Illinois (district numbers that start with “17”, displayed on the left half of the graph) than over Iowa (district numbers that start with “19”, displayed on the right half of the graph).

Figures 2.8 for corn and 2.9 for soybean show what the pixel level predictions look like. There are no true ground observations at the pixel level, so “label” maps in the top row reflect district-level ground truth observations. These maps display an intuitive climatic gradient, and built-in district borders. Our model predictions, shown in the second row, reproduce some of this gradient, and display a smoother pattern across district borders. They capture some interannual variation, like earlier corn silking in Iowa in 2007, or later soybean flowering in Illinois in 2003, but they do not capture the full magnitude of those variations, as for corn in Iowa in 2003 and 2008. Our model predictions look quite different across years, similar to the ground observations, which means that the model does not simply learn time-invariant characteristics of a place.

The prediction maps show quite a bit a local variation in crop phenology, maybe more than could be expected if there is strong spatial correlation in management choices along with weather conditions, and less of a clean climatic gradient than the true observations. Given that the model does not have any pixel-level ground observations to train against, it is perhaps not so surprising for within-district patterns to be a little disparate, though impressive that they are not more chaotic. We have tried training the model with pseudo pixel-level labels, imputed from the district-level ones. This yields smoother maps with stronger spatial correlation locally. However, this also leads the model to make more time-invariant predictions, and accordingly, this model does not perform much better than the null average model.

**Spectral band importance.** We investigate which spectral reflectance bands the model relies on, among the visible bands, which include red, blue, and green, the near infrared bands, which include two bands, and the short-wave infrared bands, which include two bands as well. We run multiple iterations of the model, with different subsets of the spectral reflectance bands in input. All models include the five additional quality, zenith, and azimuth bands, which we denote “non-spectral bands”. We compute the district-level RMSE of the model for each subset: blue and near infrared (there is evidence that corn flowering is well captured by a combination of blue and near infrared; Wu et al., 2019), visible and near infrared (this matches the spectral range of higher resolution Planet Labs data), red and near infrared (the components of NDVI), visible and short-wave infrared, visible only, and green only. We also test using spectral bands only (visible, near infrared and short-wave infrared),
Figure 2.8: Pixel-level timing of 50% corn silking, “observed” and predicted. [Top panel] Label value, inferred from district observations. [Middle panel] Prediction value. [Bottom panel] Prediction error (prediction minus label), in days. Each column corresponds to a different testing year: (from left to right) 2002, 2003, 2007, and 2008.

Figure 2.9: Pixel-level timing of 50% soybean blooming, “observed” and predicted. [Top panel] Label value, inferred from district observations. [Middle panel] Prediction value. [Bottom panel] Prediction error (prediction minus label), in days. Each column corresponds to a different testing year: (from left to right) 2002, 2003, 2007, and 2008.
without the five non-spectral bands.

We find that the model does not converge when only the green band, or the blue and near infrared bands, are provided alongside non-spectral bands. The RMSE of those models is around 200 days. All of the other models converge and have a very similar performance as our original model, but the latter, which includes all bands, remains the best performing one by a small margin. From this, we conclude that the five non-spectral bands are not crucial to model learning, although their inclusion slightly improves model performance, and that including the red band from the visible spectrum, either alongside green and blue, or alongside the near infrared bands, seems important (although we haven’t tested all possible combinations of bands within the visible triplet). Changes in model performance when dropping bands in the near infrared or in the short-wave infrared are quite small, but in the direction of decreasing performance nonetheless. Overall, this supports our initial intuition that a satellite data source with a large spectral range like MODIS is an asset for monitoring crop flowering.

2.5 Use Case 2: Predicting the Onset of Spring Flowering

2.5.1 Model

In the second use case, we predict the onset of spring flowering across the Eastern US. Figure 2.10 shows how the architecture of the core model is adapted to the spring onset task. To predict the onset of spring in a given year, we choose an input sequence, which starts on September 1\textsuperscript{st} of the previous year, and ends on August 31\textsuperscript{st} of the current year. The inclusion of the previous fall season comes from evidence that for some early-flowering species, the onset of flowering depends on autumn temperature, since vernalization – or cumulative chill – is necessary to promote early flowering (Amasino, 2004). Our model is not weather-based, but the spectral signature of the previous fall and winter seasons could reflect weather-driven phenological changes that are predictive of subsequent flowering. Combined with the 8-day frequency of the MODIS series, this choice of time period implies an input sequence of $T = 46$ time points. The output of the model is a single number, which represents the day of year when spring flowering started in a given location and year. The output of the last dilated block is connected to the final output of the model by several layers: one that collapses the time dimension to a single time point, and two that collapse the feature dimension to a single number, interspersed with a Rectified Linear Unit. For the transfer learning stage, we predict the temperature-based first bloom index from the National Phenology Network, which we describe in more detail below.
Figure 2.10: **Spring onset model architecture.** Variation of the core model architecture from figure 2.2. It highlights the structure of the input and output, application-specific blocks in yellow, and the dimension of the data throughout the network. The dotted line in brackets corresponds to the dilated block structure from the core model.

### 2.5.2 Data

#### 2.5.2.1 Ground level flowering onset

To train the model, we use ground truth observations from a long-term monitoring effort of lilac and honeysuckle phenology (Rosemartin et al., 2015). Records of first flowering for lilac correspond to the date when at least 50\% of lilac flower clusters have at least one open flower; for honeysuckle, they correspond to the date when about 5\% of honeysuckle flowers are open. This dataset has a few advantages. First, the plant phenological stage monitored follows a consistent definition over time and space. In addition, the first flowering metric directly measures onset of flowering, so there is no need to derive the timing of onset from repeated observations of flower presence and absence. Second, the dataset has been scrutinized for outliers and other quality issues. Third, it covers a decent number of years and locations. Locations are displayed as dark blue triangles on figure 2.11. After averaging observations that fall within the same MODIS pixel, there are 100 pixel-by-year data points in our restricted study region, and 991 pixel-by-year data points in the larger Eastern US region.

The main limitation of this dataset is that it is not representative of plant composition in each MODIS pixel. Lilac and honeysuckle have a phenology that is particularly weather sensitive, so their flowering onset likely reflects a weather-driven trigger of spring, rather than more complex driving forces of flowering in the local plant community. In addition, lilac and honeysuckle are non-native to many areas where they are observed. We have explored other ground truth data sources that have decent spatial coverage in the US, each of which has specific limitations. Digitized herbarium records and citizen science datasets like iNaturalist document flower presence and absence; spatial coverage comes at the expense of temporal coverage in those datasets, so they don’t have enough temporal representation in each location to allow the derivation of a reliable onset metric (Barve et al., 2020). The National Ecological Observatory Network monitors the phenology of three dominant species at each of their sites, but their data cover a few years only, whereas we want to capture interannual variations; in addition, monitoring is infrequent, which makes it difficult to derive an onset metric.
Pollen count is monitored at a number of stations across the US, but it is not publicly available across many stations. We could have combined a couple of data sources, but we chose to preserve consistency in the spring onset event definition, as additional datasets would not have increased sample size substantially.

2.5.2.2 Weather-based proxy for transfer learning

For the transfer learning step, we use the first bloom index from the National Phenology Network (Crimmins et al., 2017). This index is the output of a model that predicts the onset of flowering, using PRISM temperature and latitude-derived day length (Ault et al., 2015b,a). The model is trained with the same lilac and honeysuckle ground observations that we use to fine tune the model after transfer learning\(^\text{12}\). Essentially, the first bloom index captures the time when enough chilling days followed by enough warm days have accumulated, so that lilac and honeysuckle flowering is triggered.

2.5.2.3 Study region and summary statistics

The core study region is outlined in black in figure 2.11. We primarily train and test the model in this region, to keep memory and computational times manageable. We chose the location of the core region so that it includes a diversity of land covers and model learning does not specialize on a dominant land cover\(^\text{13}\). We then test the extrapolation skills of our model by predicting the onset of spring across all ground truth locations in the Eastern US, displayed as dark blue triangles in the figure. We exclude water and snow pixels from the study region, based on MODIS-derived land cover. The core study region outlined in black does not maximize overlap with ground truth locations, we chose it primarily to cover a large variation of land cover types.

The standard deviation of the first bloom index at each location approximates the extent of interannual variations in the onset of spring. Across the core study region, the standard deviation is 8 days on average. This means that, on average, a \(\pm 8\)-day interval around the average first bloom date is needed, to capture the first bloom date in 60\% of the years at a given pixel. There is no spatial pattern in this variability (see map in appendix B.2.4). Across the more restricted set of ground truth locations, the standard deviation of lilac and honeysuckle first flowering is 8.4 days in the core study region, and 9.2 days in the broader Eastern US. This implies that the onset of spring is quite variable,

\(^{12}\)This implies that our model might not learn much between the transfer learning stage and the fine tuning stage, except if the first bloom index does not predict lilac and honeysuckle phenology very precisely.

\(^{13}\)We did not maximize overlap of the core region with ground truth locations, since we finalized the choice of ground truth data source later on.
and that predictions should have an error strictly smaller than 8 to be informative, the lower the better. Despite the 8-day frequency of the MODIS data, we can aim for a strictly smaller error. This is because, for each pixel and each 8-day period, we know the exact date at which the “best” MODIS image selected was taken. The model can thus map spectral reflectance data to exact days, even if temporal coverage is discontinuous.

2.5.3 Results

Transfer learning: intermediate performance.

At the intermediate stage, the model is trained to predict the temperature-based first bloom index. Figure 2.12 shows the label (truth), model prediction, and model error across our study region for three randomly selected test years. The first row shows that the true first bloom index ranges between late February and early May in the region. The redder places see the earliest flowering, the bluer places see the latest flowering. The latitudinal temperature gradient is clearly visible. The second row shows that our model prediction reproduces the temperature gradient. In addition, it does not quite
Figure 2.12: **First bloom index over the core study region: observed and predicted.** [Top panel] Label value (first bloom index). [Middle panel] Prediction value. [Bottom panel] Prediction error (prediction minus label), in days. Each column corresponds to a different testing year: (from left to right) 2002, 2003, and 2009.

capture year-to-year differences: the prediction looks a lot more similar across the three test years than the true label does. The third row shows that the error ranges between -25 days and +25 days, but often falls between -10 days and +10 days.

Table 2.1 reports the RMSE of the model in each of the three test years. We compare it to the error of a null model, which uses the average bloom index value across years at each pixel to predict the outcome. Our model performs better than the null model, although the magnitude of improvement varies by year.

**Fine tuning: final performance.**

At the fine-tuning stage, we re-train the parameter values of the last fully connected layer, while
Table 2.1: Predictive performance of the model at the intermediate transfer learning stage, across the core study region. The predicted outcome is the first bloom index. The null model uses the pixel-level average index value across training years to predict the index in every year.

<table>
<thead>
<tr>
<th></th>
<th>Null model (average across train years)</th>
<th>Our model (intermediate stage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root Mean Squared Error (days)</td>
<td>7.76</td>
<td>9.17</td>
</tr>
</tbody>
</table>

keeping the model structure and all of the other parameters fixed. Table 2.2 shows that the RMSE at this stage is about 8.6 days over our study region. This is an improvement over the 9.3 days error of the first bloom index\textsuperscript{14}, and an improvement over the 9.8 days error of the intermediate stage model, but not small enough to capture detailed interannual variations in the onset of spring. However, this average error masks heterogeneity. Figure 2.13 shows a histogram of pixel-level prediction error: there are a few pixels where the model performs really badly, but the majority of pixels have an error within ±5 days.

We test the performance of the model across ground-truth locations in all of the Eastern US\textsuperscript{15}. This tells us how good our model is at extrapolating over space. Table 2.2 shows that both the temperature model and our intermediate model have a very high errors (35 days), while our fine-tuned model has a much smaller error of about 12 days. Again, this average number masks heterogeneity in model performance (see histogram in appendix B.2.5).

2.6 Discussion

Crop flowering progress. We predict crop flowering progress with good precision at the district level, and improve over both a null average model and existing models from the literature that aim for similar large-scale predictions. Our predictions extrapolate well over time, and capture well many interannual variations at the district level, though fail to capture some of the largest variations, and variations in Northern and Western Iowa.

Local pixel-level predictions are not yet reliable, as they display less spatial synchronization than

\textsuperscript{14}The first bloom index has a slight edge, since it is trained on the full set of ground truth data points, whereas our model is trained on a randomly restricted set of points, and tested on the disjoint complement.

\textsuperscript{15}To be clear, the model is still trained only on data from the core study region, but it is tested on data points outside of this region.
Table 2.2: Predictive performance of the model at the fine tuning stage, across ground truth locations. The predicted outcome is the date of lilac and honeysuckle first flowering. Performance is compared to that of the first bloom index, and of our intermediate stage model. The top portion is restricted to the core study region; the bottom portion covers the broader Eastern US.

<table>
<thead>
<tr>
<th>Study region</th>
<th>Root Mean Squared Error, days (averaged over 5 random train/test splits)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Our model (final stage)</td>
</tr>
<tr>
<td></td>
<td>First bloom index</td>
</tr>
<tr>
<td></td>
<td>Our model (intermediate)</td>
</tr>
<tr>
<td>Eastern US</td>
<td>Our model (final stage)</td>
</tr>
<tr>
<td></td>
<td>First bloom index</td>
</tr>
<tr>
<td></td>
<td>Our model (intermediate)</td>
</tr>
</tbody>
</table>

Figure 2.13: Variability in model performance across the core study region. Probability histogram of prediction error, relative to ground-truth observations of lilac and honeysuckle first flowering in the core study region. The model used is the fine tuned model.
expected given likely spatially correlated weather and management patterns. Ideally, this could be addressed by training the model on higher resolution ground truth data. This could also be addressed in the absence of such data, by adding information on spatial structure to the model, so that the model learns from clusters of pixels, and proximity between pixels better translates into correlated phenology patterns. This is a step that we plan to incorporate in future work.

The coarse spatial resolution of MODIS spectral reflectance data is not an issue for crop monitoring over the US Corn Belt, given the average field size. However, it might hinder the extension of the model to regions with intercropping or small scale farming, as each MODIS pixel will mask underlying variation in the timing of crop phenostages. Temporal resolution in satellite data has proven to compensate for coarse spatial resolution for crop yield estimation in some Indian states (Duncan et al., 2015a), so temporal frequency in satellite data is likely to remain key as well for extending our crop flowering model to such regions. In addition, these regions where intercropping and small scale farming is common, require appropriate, higher resolution, ground truth data.

The 8-day frequency of the MODIS product does not prevent the model from capturing crop flowering with an error smaller than 8 days, given that each MODIS image has a precise day-of-year timestamp. However, model predictions could likely be improved by incorporating satellite data with a higher temporal frequency, such as the daily MODIS observations of red and near infrared bands, which are available at a 250 m resolution. Both satellite data sources could be used simultaneously as inputs to the model.

We focus here on predicting crop flowering, but the USDA NASS records the timing of all stages of crop development, including planting. It would be useful to have the model predict the full sequence of crop phenostages. Beyond creating a richer set of predictions, it would likely improve predictions of flowering progress: it could help the model capture flowering shifts that are due to management-driven or regulation-driven shifts in planting date, and more generally learning multiple tasks simultaneously is a documented strategy for improving a model’s performance on individual tasks.

Our proposed remote-sensing based crop flowering proxy can be extended over a long period of time and a large geographical area. It can be integrated with other spatial datasets to identify causes of yield gaps, for example by exploring the sensitivity of crops to weather and extreme events experienced during the flowering stage. It could be extended across other major agricultural regions, and used to improve identification of large scale atmospheric teleconnections that pose global correlated risk to crop yields (Anderson et al., 2017, currently use a time invariant crop calendar to delimit the crop flowering period).
**Spring flowering onset.** We predict spring onset with greater precision than the existing temperature-based first bloom index. Our spring onset predictions also extrapolate better over space. This is an important feature. Weather-based phenological models are tied to the weather sensitivity of observed plants, which might not reflect well weather sensitivity of flowering across large and climatically-varied regions. The first bloom index model does include latitude in addition to temperature, but no other factor characterizing the local climate and environment. Models that rely on spectral reflectance data are not constrained by specific phenological responses. If spectral features are diverse enough to capture a wide range of possible spring flowering transitions, this could explain the greater extrapolation power of the satellite-based model.

In practice however, the accuracy of our spring flowering predictions varies across space, and is often, but not always, good enough to capture interannual variations in the timing of this event. This could be because the model is primarily trained on the similarly limited first bloom index, or because of the disconnect between pixel-level spectral reflectance data and non-representative phenology of cloned lilac and honeysuckle cultivars. There is a tradeoff between predicting a locally representative event, which is likely measured with error, due to data constraints, and predicting a precisely defined event like lilac and honeysuckle flowering, which could be disconnected from local patterns. Despite this limitation, our model predicts reasonably well a weather-driven trigger of spring flowering. This long-term and large-scale proxy for the onset of spring could be used to investigate phenological responses to climate change, thus complementing large data collection efforts. This is particularly important, as long-term observational data have revealed a greater sensitivity of plant flowering and leafing to temperature, compared to that suggested by experimental evidence (Wolkovich et al., 2012).

Both use cases illustrate potential benefits and limitations from using remote sensing and a weather-based transfer learning approach to monitor plant flowering. This approach is flexible and can be adapted to predicting various flowering metrics, as long as appropriate ground truth data are available.
Appendix B

Appendix to Chapter 2
B.1 Data

B.1.1 USDA NASS crop progress reports

USDA NASS publishes weekly crop progress reports, which are based on survey data collected weekly between early April and end of November. These are non-probability surveys, conducted by more than 5,000 reporters, who visit and visually assess crops across farms in their county. Reporters estimate crop progress along all stages of development, based on standard definitions. The flowering stages that we focus on are defined as follows: corn silking is “The emergence of silk like strands from the end of ears. Occurs approximately 10 days after the tassel first begins to emerge from the sheath or 2-4 days after the tassel is emerged.”, while for soybean blooming, “A plant should be considered as blooming as soon as one bloom appears.” In addition, crop progress relates to acres, and “an acre should be considered in or beyond a phenological stage when 50 percent or more of the plants in that acre are in or beyond that stage.” (USDA NASS online documentation).

Crop flowering progress, reported at the county-level, is aggregated at the agricultural statistical district level by averaging progress across reporting counties. Each county is weighted by its acreage, as estimated by NASS. The outcome is cumulative flowering progress, between 0 and 100, by crop, agricultural statistical district, and year.
B.2 Additional figures

B.2.1 USDA NASS crop flowering progress

Figure B.1: USDA NASS corn silking progress, by agricultural district (column) and year (row). The black dashed line highlights the date on which 50% flowering is reached in a given district and year. The gray dashed line highlights the average date of that event over all districts and years. Raw weekly progress data is linearly interpolated to yield a daily curve.
Figure B.2: USDA NASS soybean flowering progress, by agricultural district (column) and year (row). The black dashed line highlights the date on which 50% flowering is reached in a given district and year. The gray dashed line highlights the average date of that event over all districts and years. Raw weekly progress data is linearly interpolated to yield a daily curve.
B.2.2 Observed and predicted date of 50% flowering (district-level)

Figure B.3: District-level timing of 50% corn silking, observed and predicted. Predicted and observed dates when 50% of corn silking is reached, by agricultural district (gray label in the panel) and test year. Crosses are predictions, dots are true labels. The four test years are randomly selected.
Figure B.4: District-level timing of 50% soybean flowering, observed and predicted. Predicted and observed dates when 50% of soybean flowering is reached, by agricultural district (gray label in the panel) and test year. Crosses are predictions, dots are true labels. The four test years are randomly selected.
B.2.3 Observed and predicted crop flowering progress (district-level)

Figure B.5: Observed and predicted corn silking progress, by agricultural district (column) and test year (row). The solid lines represent observed flowering progress, linearly interpolated from weekly data. The dashed lines show predicted flowering progress, aggregated from pixel-level predictions.
Figure B.6: Observed and predicted soybean flowering progress, by agricultural district (column) and test year (row). The solid lines represent observed flowering progress, linearly interpolated from weekly data. The dashed lines show predicted flowering progress, aggregated from pixel-level predictions.
Figure B.7: Interannual variability in the first bloom index. Standard deviation of the first bloom index over 2001-2018, in days, over the core study region.
B.2.5 Spring onset model: predictive performance across the Eastern US

![Image](image.png)

Figure B.8: **Variability in model performance across the Eastern US.** Probability histogram of prediction error, relative to ground-truth observations of lilac and honeysuckle first flowering across the Eastern US. The model used is the fine tuned model.
Chapter 3

How do economic models of crop choice portray the benefits of diverse crop ecosystems?
Abstract

There is growing empirical evidence on how and when different agricultural diversification practices, benefit both productivity and biodiversity, and thus improve the sustainability of our food production systems. Diversifying crop species over space and time is a particularly salient diversification practice for its relative accessibility. In this chapter, I examine how economic models of crop diversity choice portray the economic benefits of crop diversity, and highlight potential gaps. I start by reviewing the main mechanisms through which crop diversity can benefit yield levels and stability, such as the provision of natural pest control. Building on this review, I identify several specificities of crop diversity benefits, which influence the derivation of economically optimal levels of crop diversity: these specificities relate to spatial externalities, temporal externalities, and the presence of low probability catastrophic events. I highlight how existing economic models of crop diversity choice reflect some of these specificities, and how there remains gaps between our ecological understanding and economic modeling of crop diversity benefits. This review sets the stage for improved economic modeling of crop diversity choice, which has implications for discussions of agricultural and trade policies.
3.1 Introduction

Human dominated landscapes play a key role in maintaining global diversity. In particular, they can complement and enhance the effectiveness of protected areas: they provide habitat and resources for some species, and support the dispersal of others, including for migration or climate adaptation purposes (Kremen and Merenlender, 2018). In return, some of this biodiversity directly underpins the provision of ecosystem services in productive landscapes. Services like pest and disease control, pollination, storm buffering, stable water availability, or nutrient cycling, are key for ensuring resilient and sustainable productivity of the land (Isbell et al., 2017; Garibaldi et al., 2013), and are of course dependent on how the land is managed.

In croplands, diversified farming is a salient biodiversity-based management technique for ensuring sustainable and resilient food production (Kremen and Miles, 2012; Isbell et al., 2017). However, from an economic point of view, diversity in agricultural landscapes is still often perceived as competing with high productivity, on the basis of comparative advantage and economies of scale (Weitzman, 2000; Holmes and Lee, 2012; Bellora and Bourgeon, 2019). There is a rich literature on yield benefits from crop diversity. Yet, this richness is not fully portrayed in economic models of crop diversity choice. In this chapter, I explore this potential gap. In particular, I identify and describe several specificities of crop diversity benefits, which matter for the derivation of socially optimal levels of crop diversity. I highlight how existing economic models of crop choice reflect some of these specificities, and how this literature could be expanded to further improve the characterization of crop diversity benefits.

As a clarification, diversified farming encompasses agricultural production systems that incorporate agrobiodiversity across multiple ecological, spatial and temporal scales (Kremen and Miles, 2012, following Altieri, 2004). I focus here on crop diversity practices, such as mixed crop plantings or rotating crops, as they are promising and accessible techniques for increasing productivity without chemical control (Kremen and Miles, 2012). More precisely, I consider crop diversity across multiple spatial and temporal scales, from within-field intercropping, to field-level crop rotations or cover cropping, to cross-field crop mosaics. I focus on diversity in crop species, rather than in crop variety within species, genetic diversity within varieties, or diversity across crop and non-crop plantings.

The rest of the chapter is organized as follows. Section 3.2 reviews the main channels through which crop diversity in agricultural systems influences the level and stability of crop yield. Section 3.3 outlines potential sources of disconnect between our ecological understanding and economic treatment of crop diversity benefits. Specifically, I discuss how spatial externalities, temporal externalities, and objective
functions that weigh in low probability catastrophic events, influence the derivation of optimal crop diversity levels. I explain how existing economic models of crop choice address some, but not all of these specificities. Finally, section 3.4 illustrates one particular implication of spatial externalities: they can create path dependency in crop diversity choice. I model crop diversity choice across a grid of individually-managed crop fields, and model arthropod-driven pest control across this grid. I show that, under some conditions, multiple equilibria can arise, at both ends of the crop diversity spectrum: a single grid cell will be cultivated as low diversity if the surrounding crop grid is low diversity, and high diversity if the surrounding crop grid is high diversity. This is despite the fact that under these conditions, the socially optimal grid is the highest diversity one. This highlights a particular coordination failure, which arises from the spatial externality of crop diversity benefits.

3.2 Background: various channels link crop diversity with increased crop yield, stability, and resilience

Crop diversity can benefit crop yield through multiple channels. I describe the most prominent ones in this section, and summarize some of their characteristics in table 3.2: whether the benefit is private (benefit to the farmer planting the field) or has a spatial externality (benefit to other farmers); the spatial scale of the impact (local denotes the level of a field, landscape includes the surrounding fields, and is loosely employed to represent the relevant spatial scale of pest, pathogen, or pollinator population dynamics, regional denotes larger scales, up to the scale of continents, and global denotes a worldwide scale); the temporal scale of the impact (short-term denotes days to years, long-term denotes decades or more); the human input used as a substitute in low diversity agricultural systems; the relevant crop outcome to consider, whether it is yield level, stability, or resilience to shocks; and the relevant dimension of crop species diversity to consider, whether it is spatial or temporal. The rest of this section documents each channel in more detail: for each channel, it reviews evidence from the literature, and uses it to derive the characteristics documented in table 3.2.

Pest control. To start with, higher crop diversity usually implies lower pest pressure. Larsen and Noack (2017, 2021) find that crop diversity decreases both the level and the variance of insecticide use, even when taking into account average differences in insecticide use between crop types and farmers, or changes in insecticide use due to regional weather and policy shocks. This pattern arises from a variety of mechanisms.
Table 3.1: Key channels linking crop diversity and crop productivity, along with the relevant spatial and temporal scales and other characteristics.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Private or externality</th>
<th>Spatial scale</th>
<th>Temporal scale</th>
<th>Human substitute</th>
<th>Crop outcome</th>
<th>Diversity dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>lower pest abundance</td>
<td>externality</td>
<td>local to landscape</td>
<td>short-term</td>
<td>higher yield</td>
<td>spatial + temporal</td>
<td></td>
</tr>
<tr>
<td>lower damage per pest (lower probability of evolving a harmful genetic combination)</td>
<td>externality</td>
<td>landscape / regional / global</td>
<td>long-term (evolution)</td>
<td>pesticide</td>
<td>spatial + temporal</td>
<td></td>
</tr>
<tr>
<td>lower damage per pest (distraction by other, less susceptible plants)</td>
<td>private</td>
<td>local</td>
<td>short-term</td>
<td>higher yield</td>
<td>spatial + temporal</td>
<td></td>
</tr>
<tr>
<td>interruption of disease cycle</td>
<td>private</td>
<td>local</td>
<td>short-term</td>
<td>fungicide, bactericide</td>
<td>temporal</td>
<td></td>
</tr>
<tr>
<td>lower disease transmission</td>
<td>externality</td>
<td>local</td>
<td>short-term</td>
<td>higher yield</td>
<td>spatial + temporal</td>
<td></td>
</tr>
<tr>
<td>lower disease transmission</td>
<td>externality</td>
<td>landscape / regional / global</td>
<td>long-term (evolution)</td>
<td>higher yield</td>
<td>spatial + temporal</td>
<td></td>
</tr>
<tr>
<td>higher pollinators abundance and diversity</td>
<td>externality</td>
<td>local</td>
<td>short-term</td>
<td>imported farmed bees</td>
<td>higher yield</td>
<td></td>
</tr>
<tr>
<td>niche partitioning (higher resource use)</td>
<td>private</td>
<td>local</td>
<td>short-term</td>
<td>spatial + temporal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>higher nutrient availability</td>
<td>private (externality in the long-run)</td>
<td>local</td>
<td>short-term + long-term (soil health)</td>
<td>fertilizer</td>
<td>higher yield</td>
<td></td>
</tr>
<tr>
<td>higher water regulation</td>
<td>private</td>
<td>local</td>
<td>short-term + long-term (soil health)</td>
<td>irrigation</td>
<td>higher yield</td>
<td></td>
</tr>
<tr>
<td>insurance hypothesis</td>
<td>externality</td>
<td>regional</td>
<td>long-term</td>
<td>trade and insurance programs</td>
<td>higher yield stability</td>
<td></td>
</tr>
</tbody>
</table>

First, there is evidence that more diverse fields support lower pest abundance. In particular, a long line of research has documented evidence of the resource concentration hypothesis, which posits that “herbivores are more likely to find and remain on hosts that are growing in dense or nearly pure stands” (Root, 1973). This hypothesis implies a dilution effect of diverse crop stands. Wetzel et al. (2016) contributes an interesting mechanism: they find that bugs surrounded by diverse plants are harmed much more by low-quality plants with the wrong nutrient levels than they are benefited by high-quality plants with high nutrient levels; conversely, homogenous farm fields create monocultures where pests may find the perfectly homogenous nutrition to be healthy and reproduce. This is reminiscent of the dilution effect described in Ostfeld and Keesing (2000), where species diversity among tick hosts can result – though not always – in lower risk of human exposure to Lyme disease, as more ticks feed on inefficient disease reservoirs. In the case of Wetzel et al. (2016), it is instead plant diversity that
dilutes the efficiency of nutrition for pests.

Diversifying production systems with mixed crops or crop rotations not only reduces pest colonization and population growth rates by diluting resources for herbivores, it also increases the abundance and diversity of their natural enemies (Isbell et al., 2017, and review by Kremen and Miles, 2012). In the 1950s, Elton observed a higher probability of pest outbreaks in agriculture monocultures, and proposed that higher plant diversity stabilized animal food web dynamics, in particular, maintaining natural pest enemies (Elton et al., 1958). Landis et al. (2000) review habitat management strategies that provide biological control of arthropod pests by conserving their natural enemies. Much has to do with spatial and temporal diversity of the broader agricultural habitat, which includes non-crop elements, but crop diversity matters too: cotton-wheat intercropping systems implemented in China successfully reduce cotton aphid pressure and increase productivity by supporting cotton aphid predators; suggested habitat management strategies also include taking into account adjacent crops when planning crop rotations. Fahrig et al. (2015) find a negative correlation between crop landscape homogeneity, captured by average crop field size, and the diversity of natural pest enemies across multiple taxa. They hypothesize that this is due to greater access to field boundary habitat in more heterogenous crop landscapes.

In addition to resource dilution and support of natural enemies, crop diversity can lower pest abundance by hampering the movement and colonization of pests across crop patches (Risch et al., 1983; Margosian et al., 2009; Meehan et al., 2011). This is because a diverse crop landscape is less connected than a monoculture landscape, from the point of view of a specialist crop pest. This is true for spatial crop diversity but also temporal crop diversity, depending on the frequency of crop turnover relative to the pest’s generation time. Theoretical work has shown that temporal variability in a landscape pattern can impact populations even more strongly than spatial variability (Fahrig, 1992; Hodgson et al., 2009).

Overall, this pest abundance mechanism has clear spatial externalities, as the nutritional base of pest populations and control exerted by its natural enemies are likely to extend beyond field boundaries. This spatial externality is likely to be most relevant at small scales. Control of arthropod pests for example is studied at the local scale and at the landscape scale (Kremen and Miles, 2012). Landis et al. (2008) suggest that habitat diversity up to 1.5 km away from a soybean field decreases soybean aphid abundance in that field, while Larsen and Noack (2021) find that crop diversity up to 2 km away from the focal field impacts local insecticide use. Both spatial and temporal dimensions of crop diversity matter for this pest abundance channel.
Second, more diverse crop fields can curtail the damage potential of pests, by slowing down the evolution of a deadly genetic combination targeting individual crops. Alstad and Andow (1995) simulate how resistance of the European corn borer to transgenic maize can evolve more slowly in fields mixing toxic and non-toxic maize than in monotoxic maize fields—in 20 versus 2 generations respectively, which translates into 5 versus 1 year, assuming an average 2 generations per year. This slow down occurs thanks to gene exchange between selected and unselected insect populations from adjacent toxic and non-toxic maize fields. Years to decades seems like a good approximation for the temporal scale of this effect, as generation times for crop pests usually range between 5 days and one year, with many values falling around 30-100 days, and some insect pests develop resistance to chemical insecticides within 50 generations of first exposure (Heimpel et al., 2005). The extent of the spatial externality is likely local in the short-term, but spreads out to a global scale in the long-term. Both spatial and temporal dimensions of crop diversity matter for this pest damage potential channel.

Third, more diverse crop fields can curtail the damage potential of pests through diversion: in well-designed mixed plots, some crops can serve as a diversion to decrease pest damage on more susceptible crops (Altieri, 2004). This might be more relevant for combinations of crops with non-crop plantings, smaller-scale gardens, or closely intertwined crops, although it is exploited by planting alfalfa on the perimeter of cotton in California to reduce damage from cotton bug (Wenda-Piesik and Piesik, 2021). This is a short-term and localized effect.

All three channels linking crop diversity and pests ultimately impact yield levels and the probability of pest or pathogen driven widespread crop failure. The human substitute for this diversity-mediated natural pest control is the application of pesticides (fungicides, herbicides, and insecticides).

**Pathogen control.** The impact of crop diversity on microbial pathogens is less predictable than the impact of crop diversity on pests, due to the strong influence of the microclimate (Lin, 2011). Nevertheless, crop diversity can provide effective pathogen control through various mechanisms. Temporal diversity in crop rotation can suppress crop diseases, by interrupting disease cycles. This applies to pathogens that have a selective host range: for example, broadleaf crops are not affected by common cereal pathogens, so introducing them in a rotation for a sufficient amount of time can help break a cereal disease cycle (Krupinsky et al., 2002). Spatial and temporal diversity in polycultures can also suppress crop diseases by decreasing disease transmission, via decreased host abundance (Mitchell et al., 2002).

In addition, the same evolutionary channel mentioned for pests applies to pathogens: higher diver-
sity can impede the evolution of deadly genetic combinations. Overall, the pathogen channel primarily occurs at the local to landscape scale, in the short-term, and impacts crop yield levels. Temporal crop diversity is most relevant for the disease cycle interruption mechanism, while both spatial and temporal diversity matter for the disease transmission mechanism. Humans compensate for diversity-mediated biocontrol using relevant pesticides like fungicide and bactericide.

Pollination. Another channel through which crop diversity impacts crop yield is through the provision of pollination services. It is clear that distance from florally diverse habitat reduces pollination in crop fields (see for example Garibaldi et al., 2011; Albrecht et al., 2020, and more studies reviewed in Garibaldi et al. 2020). By extension, crop diversity could support greater pollinator communities, by providing a greater variety of floral resources over time. More generally, landscapes with higher crop diversity can support greater pollinator densities by providing more diverse food resources, and habitats for nesting and dispersal (Fahrig et al., 2011). However, experimental evidence linking crop diversity with pollination benefits is limited. Some studies find a positive impact of crop diversity on pollinator abundance or diversity (Aguilera et al., 2020; Raderschall et al., 2021; Fahrig et al., 2015), others a negative one (Hass et al., 2018; Martin et al., 2020). Results depend on scale, whether small (Martin et al., 2020), or large (Fahrig et al., 2015). They also depend on co-occurring factors. For example, Hass et al. (2018) hypothesize that the negative association that they find between crop diversity and bee abundance, might be due to crop diversity being positively correlated with intensive crop management in their study region. In addition, a recent study shows that crop diversity benefits pollinator diversity, particularly in landscapes that are rich in semi-natural habitats (Aguilera et al., 2020). Diversifying crop types at the landscape scale could thus benefit pollination by enhancing the benefits of setting aside natural habitat within cropland. There is a more consistent positive impact of smaller crop fields and perennial relative to annual crops, two features that are intertwined with crop diversity choices, on pollinator diversity (Fahrig et al., 2015; Martin et al., 2020). There remains gaps in our understanding of how impacts of crop diversity on pollinator abundance and diversity translate into pollination-mediated yield benefits. Raderschall et al. (2021) contribute to this gap, by showing that landscape crop diversity increases bumble bee densities in faba bean, and in turn insect pollination increases faba bean yield.

Crop diversity can also enhance pollination benefits by improving temporal continuity of flower resources, for pollinators with a generalist diet. Nicholson et al. (2020) find that raspberry farms that also grow earlier-blooming blueberries, have higher populations of *Bombus impatiens*, a generalist crop
pollinator, relative to raspberry farms not growing blueberries, but do not see differences in populations of specialist wild bee species.

Overall, pollination benefits of crop diversity are present but not fully understood. They occur across a similar spatial scale as pest control benefits, that is to say at the landscape level (Raderschall et al., 2021). Both spatial and temporal crop diversity influence pollination benefits, and ultimately impact crop yields. The human substitute for this diversity-mediated pollination service is farmed bees and the associated bee shipping business.

**Nutrients.** Crop diversity also impacts nutrient uptake and availability. One central mechanism is the complementarity of resource use across species, which allows resource partitioning and facilitation, and ultimately enhances resource use. For example, Loreau and Hector (2001) find that plant diversity improves primary production through resource partition and positive interactions between species, based on comparisons of grassland species performance in mixture versus in monoculture. Resource facilitation can take the form of increased nutrient availability. For example, mixing crop groups—like nitrogen fixers with non nitrogen fixers—in crop rotations increases yield through nitrogen availability (Davis et al., 2012).

Tamburini et al. (2020) perform a meta-analysis of meta-analyses looking at the impact of different diversification practices on ecosystem services. They find that crop diversity consistently improves soil fertility and nutrient cycling, as measured by increased nutrient availability, limited nutrient leaching and runoff, increased soil organic carbon, and enhanced soil bacteria and arbuscular mycorrhizal fungi. They also find that improvements in soil fertility and nutrient cycling consistently occur in tandem with improvements in crop yield. Recent work provides further insight into the fungi channel: it finds that crop diversity promotes richer and more diverse communities of arbuscular mycorrhizal fungi (after controlling for co-occurring differences in soil properties), thus countering the negative effect of agricultural intensification on these communities (Guzman et al., 2021).

Crop diversity impacts on nutrient use efficiency and availability operate locally in the short-term. With time, larger-scale nutrient cycling generates spatial externalities. As the previous examples illustrate, temporal crop diversity – or fine-scale spatial diversity – is the relevant metric. The human substitute for this ecosystem service is fertilizer use.

**Water.** The impact of crop diversity on crop yield via water quality and quantity is connected to the nutrient channel to some extent. The meta-analysis performed by Tamburini et al. (2020) finds
that crop diversity consistently improves water quality, primarily by limiting nitrogen leaching loss, and water quantity, by enhancing soil water storage. They also find that these water improvements co-occur with crop yield increases.

Crop yield stability and resilience to extreme conditions is a very relevant outcome for this water channel. Using an 11-year study in Minnesota grasslands, Tilman et al. (1994) show that high plant species richness increases resistance to and recovery from major droughts, in terms of primary productivity. The temporal dimension here is key in terms of the timeline of impacts, but temporality also matters in terms of measuring crop diversity. Bowles et al. (2020) provide evidence that long-term crop rotational diversity in maize cropping systems improves maize yield resilience to adverse growing conditions, including droughts. Crop rotation diversity, which is here guided by local practices, lowers the risk of crop failure under water stress, increases the probability of a bumper crop under good environmental conditions, and accelerates yield growth over time. Various mechanisms could explain this greater stress resistance: increased soil water storage capacity due to higher soil organic matter content and porosity, higher water retention and infiltration, changes in soil physical structure, rooting characteristics, or microbial communities.

Crop diversity benefits through the water channel primarily occur locally, in the short and in the long-term, and through temporal crop diversity (spatial diversity matters too, but to a lesser extent). Human substitutes for this water-mediated crop diversity benefit include irrigation, and farmer insurance programs.

Insurance hypothesis. There is a massive literature addressing the question of whether diversity in ecosystems begets stability. Although phrased as a debate, it reads more like a discussion, where different choices of definition for diversity and stability, different assumptions about ecosystem equilibrium, or different time horizons for empirical work, yield different relationships between the two (see Tilman et al., 1994, for some seminal references, and Ives and Carpenter, 2007, for a review of this multifaceted question). Here I am specifically interested in the association between crop diversity and crop yield stability and resilience to shocks. Locally, crop diversity provides ecosystem function benefits, which directly translate into a greater ability of the ecosystem to withstand environmental fluctuations and shocks. For instance, as mentioned in the previous paragraphs, crop diversity supports natural enemy populations, which constitute a buffer against climate-mediated pest outbreaks; crop diversity improves soil water storage capacity, which helps withstand fluctuations in water supply. Such functional channels can explain field-level empirical evidence of yield stability increasing with
temporal crop diversity (Snapp et al., 2010; Gaudin et al., 2015).

At larger temporal and spatial scales, the insurance hypothesis – or portfolio effect – plays an important role. The idea is that, under some conditions, diversity provides insurance by buffering against environmental fluctuations, because different species respond differently to those fluctuations, such that ecosystem function is maintained (Naeem and Li, 1997; Yachi and Loreau, 1999). This can lead to enhanced performance on average over the long-run (Yachi and Loreau, 1999; Loreau et al., 2003). Renard and Tilman (2019) test this insurance hypothesis empirically at the national scale. They find that both higher crop species diversity and higher crop group diversity are associated with lower standard deviation of national yield, after controlling for climate and external inputs. In addition, average national yield is statistically independent of crop species diversity, but positively associated with crop group diversity, which could be due to the nitrogen fertilization effect of legumes. Abson et al. (2013) also find a stabilizing affect of agricultural diversity at the landscape level.

Decades of research in ecology and related fields thus add to millennia of knowledge accumulation in agroecology, to document the extent of benefits of crop diversity. I focus here on benefits that directly impact economic productivity. They occur through diverse channels, many of which have spatial externalities, especially at the landscape scale. Some benefits occur over the long term, and temporal crop diversity is at least as important as spatial crop diversity. In the next section, I highlight how some of those characteristics call for careful incorporation into economic models of optimal crop diversity, and review how existing models portray the benefits of crop diversity.

3.3 Potential sources of disconnect between ecological characteristics and economic modeling

3.3.1 Key characteristics of crop diversity benefits

3.3.1.1 Spatial feature: large scale spillover of crop diversity benefits

As shown in table 3.2, many crop diversity benefits involve spatial externalities, primarily the pest, pathogen, pollination, and insurance channels. The extent of the externality varies, but many occur at the local to landscape scale, although the scale increases when longer timescales are considered,

1They find that agricultural diversity decreases both volatility in returns and average returns. However, their data on returns are UK-wide and only disaggregated by agricultural activity, so they could be missing landscape-level and field-level yield enhancements from agricultural diversity.
and the insurance channel occurs at all scales. As with all externalities, this leads to a disconnect between private farmers’ decisions about crop diversity on their land, and the socially optimum level of crop diversity. Specifically, not internalizing the positive spillover of crop diversity to neighboring crop fields leads to underinvestment in local crop diversity, or suboptimal crop diversity at larger scales, due to lack of coordination among farmers. This has often been incorporated into economic models of crop diversity choice, although the spatial scale chosen to model the externality does not always match empirical evidence.

3.3.1.2 Spatial feature: large scale negative externality from substituting crop diversity benefits with human input substitutes

The drive towards agricultural simplification has gone hand in hand with increased reliance on external inputs like pesticide or fertilizer, to compensate for lost crop diversity services. This is manifest in the lower insecticide use observed under high crop diversity (Larsen and Noack, 2021), or in the greater yield gains from diversification, in the absence of external inputs (Smith et al., 2008). These external inputs have large negative externalities: increased infant mortality from pesticide use (Frank, 2017), greenhouse gas emissions, air, and water pollution from Nitrogen fertilizer use (Keeler et al., 2016), harmful impact of managed honey bees on their wild relatives (Geldmann and González-Varo, 2018), to list a few.

Choice of crop type and crop diversity thus involves another type of spatial externality: substitution of crop diversity benefits by external inputs imposes spatially diffuse environmental and health costs. Economic models usually consider market prices as the cost of external inputs, which is strictly lower than their social cost and biases the productivity-diversity tradeoff towards simplified agricultural landscapes. To be fair, the rigorous estimation of these negative externalities is a challenging task, but their existence is often omitted altogether from the discussion of optimal crop diversity levels.

3.3.1.3 Temporal feature: lagged private benefits

It is possible that the full extent of private crop diversity benefits will only fully appear at larger temporal scales than the ones usually addressed in empirical work or considered in theoretical models. This is because some benefits take place over the medium to long-term, in particular the soil health benefits that underlie the nutrient and water channels (table 3.2). Many dimensions of soil quality change slowly, as illustrated by experimental studies of crop rotation impacts on soil quality, which rest on decades of data collection (Karlen et al., 2006, 2013). In addition, the benefits that consist
in a lower probability of extreme stress conditions (such as the occurrence of a pest outbreak, or the evolution of highly detrimental pest or pathogen genotype), and the benefits that consist in a more stable response to environmental fluctuations, can take years or more to manifest, depending on the frequency of those extreme stress events or fluctuations.

This feature is directly relevant for empirical estimates of private crop diversity benefits\(^2\), which are not my main focus. It does also matter for models of crop diversity choice, as farmers can be assumed to be more or less myopic in their perception of diversity benefits, and empirical estimates impact model calibration.

3.3.1.4 Temporal feature: temporal (and spatial) externality via evolutionary forces

Crop diversity benefits generate temporal externalities. This is because some consequences of today’s crop choice, which manifest in the future, do not necessarily incur to today’s farmer, if there is high turnover in land management or if farmers plan for a short time horizon. Lagged soil impacts described in the previous section could generate such an externality, if lags are not internalized by farmers choosing their crop mix. A temporal externality could also arise from the impact of today’s crop choice on the temporal sequence of crops grown at that location, as this temporal sequence has uncertain and potentially distant consequences on the evolution of harmful pest and pathogen strategies.

The logic is as follows: the crop mix chosen today either contributes to a steady pattern of monoculture or simple rotation, or it diversifies this pattern by introducing new crops or changing the frequency or complexity of a crop rotation. Temporal diversity, in turn, impacts the evolutionary adaptation of pests and pathogens (table 3.2). Crop enemies are more likely to evolve a deadly genetic combination targeting a specific crop if that crop is the unique and stable adversary (hence gene selection is consistent over time), rather than if crop types vary over time. As mentioned in section 3.2, this has been illustrated with a simulation in the case of spatial crop diversity: mixing transgenic maize with non-transgenic maize can delay the evolution of resistance to transgenic maize among the European corn borer, because gene selection differs between transgenic and non-transgenic patches, and genes mix across those patches in the landscape-level insect population (Alstad and Andow, 1995).

\(^2\)Empirical estimates that rely on repeated observations of crop yield at a few sites, where crop diversity changes over time, require long time series to allow soil quality changes to fully materialize. Empirical estimates that rely on cross-sectional comparisons of crop yield across a crop diversity gradient need to control for other differences along this crop diversity gradient, which affect crop yield. This is very difficult to achieve, as farmers take into account factors like soil quality and environmental fluctuations when choosing which crops to grow and the planting pattern. Across both types of estimates, different results are likely to be obtained, depending on the severity of stress events or environmental fluctuations covered in the data sample.
It is possible that farmers internalize this temporal externality if they are relatively far-sighted, and the temporal externality relatively short-term. However, they are unlikely to internalize it fully, as there is unpredictability, and spatial externalities contribute to the dynamics described. This implies that the privately optimal choice of crop diversity is likely to achieve a lower level of temporal crop diversity than what is socially optimal. I have not seen the existence and implications of such temporal externalities addressed in economic models of crop diversity choice.

3.3.1.5 Low probability catastrophic events

One particular benefit of crop diversity, is a decreased risk of low-probability catastrophic crop failure from a pest or pathogen attack. This type of risk is present in other domains, such as climate change, and has been analyzed theoretically. In particular, it has been shown that for certain classic types of utility function (the welfare metric that the economic model tries to maximize), low probability catastrophic events can either insufficiently or excessively influence the choice of the optimal policy function, which here translates into the choice of the optimal crop diversity level.

To start, policy choice under uncertainty can be insensitive to low probability catastrophic events. Chichilnisky (2000) show that, under the assumption of bounded utility, which precludes extremely low welfare levels under catastrophic events, policy choice under an expected utility framework is insensitive to small probability events. They argue that this does not match how humans actually treat choices under uncertainty, based on experimental evidence. They propose a new objective function, which combines an expected utility term and a term that directly reflects avoidance of catastrophic events. This proposed objective function is sensitive to both low probability and large probability events.

Weitzman (2009, 2011, 2014) agree that policy choice should depend on the treatment of low-probability catastrophic events. They use a simplified model to show how this can occur: under some specific conditions, which include sufficiently low consumption levels under the low probability catastrophic scenario\(^3\), policy choice can be sensitive to low probability catastrophic events, to the point that these events dominate economic calculations\(^4\). They then make the point that studies usually don’t take fat-tail events into account, so that standard cost-benefit analysis is actually impervious to uncertain probabilities and consequences of catastrophic events.

\(^3\)This is incompatible with the assumptions from Chichilnisky (2000).

\(^4\)They make this argument about the social cost of carbon, and I translate it to the question of crop diversity: a very low probability catastrophic event can make the social cost of low crop diversity infinitely high, if catastrophic consumption levels are sufficiently low in particular, so that the consumption nudge from increasing crop diversity today moves us away from the fat tail and makes this increase worth it.
Taken together, both of these papers say that with a utility function bounded from below, we get insensitivity to low probability catastrophic events, and with a utility function unbounded from below, we can get complete dominance of low probability catastrophic events in policy choice. Both argue that some sensitivity is reasonable. This calls for careful modeling of crop diversity choice, if we believe that this choice should be sensitive to the probability of widespread catastrophic crop failure.

3.3.2 Economic modeling of crop diversity choice

In this section, I review the main economic models of crop diversity choice. I highlight their strengths and limitations, in light of the key characteristics of crop diversity benefits discussed above. All four models consider the spatial dimension of crop diversity.

3.3.2.1 Theoretical model: optimal crop diversity level, assuming an explicit pathogen channel that has global spillovers

Weitzman (2000) model the optimal choice of crop diversity, given that crop diversity provides natural pathogen control. The central ecological assumption is that the higher the biomass of a crop, the greater the number of parasite strains that specifically target this crop, which increases the probability of crop failure. In addition, this pathogen effect is assumed to create a global externality, as pathogens circulate throughout all fields (space is not explicit in the model). The authors model the social objective as a minimization of the risk of generalized crop failure, under the constraint of achieving a certain level of economic welfare. They show that this objective can be rewritten as a tradeoff between net economic profitability and a Shannon-index type of ecological diversity. The associated socially optimal level of crop diversity can be achieved by a tax or subsidy, but the model does not yield testable predictions. One policy conclusion is that the socially optimal crop mix involves a strictly positive level of biomass for every crop variety, although this prediction is tied to the choice of functional form for the pathogen effect.

The biggest strength of this model is to put forward an important spatial externality associated with the choice of crop diversity, based on the evolution of crop parasite strains, which causes the socially optimal level of crop diversity to be higher than the privately optimal level. In addition, it is interesting that the objective is modeled as a direct minimization of the low probability, high impact risk of widespread crop failure. This allows the probability of widespread crop failure to influence the choice of crop diversity, as discussed in section 3.3.1.5. Some limitations lie in the details of the central ecological assumption. Modeling pathogen pressure using a species-area relationship
lacks ecological backing. In addition, space is not explicit in the model, so crop diversity is crudely captured by crop biomass and independent of spatial patterns, and the scale of the spatial externality is essential infinite. More importantly, to model the central assumption of crop failure decreasing with crop diversity through lower pathogen pressure, it could be more relevant to consider temporal crop diversity rather than spatial crop diversity, as discussed in section 3.3.1.4. Overall, this is an interesting exposition model, but many approximations in how the pathogen channel is modeled, restrict its usefulness for bringing this model to data or further digging into how pathogen pressure actually shapes crop diversity choices.

Temporal crop diversity can easily be added to the model. Pathogen pressure is a function of cumulative crop biomass over space, and it can be modeled in addition as a function of cumulative crop biomass over time. I incorporated this change into the model, and derived the new tradeoff between economic profitability and ecological diversity. With this change, the ecological diversity term can be decomposed into (1) a Shannon-index type, like in the original model, which implies that crop proportions will be kept small, and that there will be spatial diversity, and (2) a term that interacts current crop biomass with cumulative biomass grown in the past (separately for each crop type), which implies that a higher proportion of a given crop type will be grown in the patches where it hasn’t been grown often in the past, i.e. that there will be temporal diversity. The socially optimal level of crop diversity cannot be derived explicitly from this model, which limits explorations of how different modeling choices for the spatial and temporal externalities lead to different crop diversity optima.

3.3.2.2 Theoretical model: optimal crop diversity level and tax on pesticides, assuming an implicit pest channel that has regional spillovers, and negative externalities from pesticide use

Bellora and Bourgeon (2019) model crop diversity choice at the national level, and how it is influenced by trade and taxes on pesticides. They start from a Ricardian trade model, where two countries choose which crops to grow. Under classic conditions, each country specializes in the crops in which it has a comparative advantage. They introduce crop production externalities, such that there is incomplete specialization relative to the classic setup. Specifically, they incorporate Weitzman’s story of crop diversity lowering pest pressure, and add a role of pesticide use: the survival probability of a farmer’s crop decreases with the share of the country’s cropland devoted to that crop (pest pressure effect with spatial externality), it increases with the farmer’s pesticide use, and it increases with the other farmers’
pesticide use (pesticide spatial externality). In addition, countries have a preference for pesticide use in foreign countries, due to negative pesticide impacts on environmental and human health (“not in my backyard”). They derive the optimal pesticide tax level, and implications for national gains from trade and crop production volatility.

Overall, the focus is more on the optimal pesticide tax than on the optimal level of crop diversity. Pest effects are used as a convenient way to induce decreasing returns to scale in farming, and “not in my backyard” considerations, into a trade model. Their modeling of crop diversity share similar limitations with Weitzman’s: crop diversity is crudely captured by crop shares, and the pest externality has a country-wide scale. The functional form for the crop survival probability function is set in a way that the optimal pesticide level for a farmer does not depend on other farmers’ behavior, or on local or global crop diversity. It depends only on pesticide price, on farming costs (labor wage and land rent), and on the unregulated level of pesticides. Due to the complexity of this type of trade model, it is unclear if more nuance can be incorporated to the pest pressures story (smaller spatial externality, role for temporal diversity, avoidance of crop failure into the objective function, etc.) while maintaining a tractable model.

3.3.2.3 Theoretical model, with an empirical estimation: impact of neighboring crop diversity on productivity, through an implicit pest channel

The goal of Bellora et al. (2017) is to estimate the impact of spatial crop diversity on crop productivity. They do so using satellite-derived data on South African farms. They first craft a model of crop production, where crop diversity impacts crop survival probability via a pest pressure channel. This complex modeling is flattened out in the reduced form model, where crop survival at the field-level is a log-linear function of a list of field-level characteristics, including local crop diversity. The interesting feature is the choice of spatial scale: crop diversity is measured locally, and at different radii around the focal field. The overall estimated impact of crop diversity on productivity reflects a combination of channels operating at this scale, it does not isolate the pest channel.

They find that crop diversity up to 2 km away from a focal field has a positive impact on crop productivity. They cannot control for field-level pesticide or other input use (except for irrigation), nor for soil characteristics, whereas those are likely to be correlated with both crop diversity and crop productivity. If crop diversity is associated with lower input use or lower soil quality, then the impact of crop diversity on productivity will be underestimated. In addition, this is not a panel dataset but farm and crop fixed effects are included, so the crop diversity effect is estimated from variations in
crop diversity across fields within a farm, which might not be substantial. Finally, the impact of crop diversity on productivity might be underestimated because the dataset is a snapshot in time and does not include extreme environmental fluctuations or pest events.

3.3.2.4 Theoretical model, with an empirical estimation: impact of in situ crop diversity on productivity, through an explicit nutrient channel

Noack and Quaas (in development) estimate the private value of spatial crop diversity on Ugandan farms. The hypothesized mechanism is the nutrient channel, where spatial crop diversity allows for more efficient use of in situ resources. They find a positive impact of crop diversity on farm revenues, which does seem to stem from a positive impact of crop diversity on crop yield through more efficient resource use. They then calibrate a model of the farmers’ tradeoff between exploiting economies of scale through specialization, and exploiting ecological benefits through diversification. In this model, external inputs act as substitutes to ecological benefits, though only their market cost is considered, not their negative externalities. This model predicts that easier access to external inputs will drive down the level of crop diversity chosen by farmers.

Their modeling choices are coherent with their choice to focus on the private nutrient channel. Only some of the private benefits of crop diversity are captured since temporal diversity is not considered. In addition, the sample spans about 5 years, which might not cover much environmental fluctuations.

My focus here is on how crop diversity benefits are portrayed in economic models of crop diversity choice. This helps understand which benefits are reflected in derivations of private and socially optimal levels of crop diversity. Hence I have surveyed theoretical models of crop diversity choice, some of which are paired with empirical estimates of crop diversity impacts on productivity. There are more economic papers around this topic, which focus on the empirical estimation of crop diversity benefits, such as the private benefit of spatial crop diversity for poverty (Di Falco et al., 2010, Ethiopia), for production level and variance (Michler and Josephson, 2017, Ethiopia), for consumption (Tefsaye and Tirivayi, 2020, Uganda); the insurance benefit of crop diversity (Auffhammer and Carleton, 2018, at the district level, Renard and Tilman, 2019, at the national level); or the pest control externality (Larsen and Noack, 2017, 2021). Those papers complement and sometimes have been mentioned along the rich ecology literature surveyed in the first section of this chapter, on crop diversity benefits for crop yield. There

---

5Crop diversity could also allow more efficient use of farm labor resources, not just environmental resources, but the evidence they find is in support of the latter.
are also economic papers modeling crop diversity choice as a toy illustration of a broader economic concept. Those papers tend to abstract from the crop diversity benefits that I outline in this chapter. For instance, Holmes and Lee (2012) model and derive how crop choice depends on local versus adjacent soil quality, to estimate how much crop specialization comes from natural advantage (choice driven by local soil quality) versus economies of scale (choice driven by adjacent soil quality, which drives adjacent crop choice); they assume no additional spillover impacts from neighboring fields, so no pest channel.

Overall, there are interesting theoretical explorations and empirical estimates of some of the economic benefits of crop diversity. On the theory side, more benefit channels could be explored, with models tailored to the particular spatial and temporal features of each benefit, and care taken in the design of the social planner’s optimization objective. The role of temporal crop diversity in particular, and the externalities it generates, is underrepresented in this literature, as far as I have seen. In the next section, I consider the role of crop diversity as supporting populations of natural pest enemies, and expose how path dependence can arise from this spatial externality.

### 3.4 Illustration: neighborhood effects and multiple equilibria

In this section, I come back to one crop diversity benefit, arthropod mediated pest control, which generates a spatial externality at the landscape level. I use a simple model of crop choice across a grid of crop fields, to illustrate some consequences of such an externality for private and socially optimal levels of crop diversity. In particular, I show how private decisions of crop diversity can be influenced by diversity in the surrounding crop fields: a low diversity equilibrium can be sustained even when the high diversity equilibrium is socially optimal.

I consider a grid of \( N \) fields, where each field has the same size, and is managed by a different farmer. The entire grid spans about 1.5 km\(^2\), to match a typical scale of arthropod population dynamics. A farmer managing field \( i \) chooses between growing the dominant monoculture in this landscape (\( i = m \)) and growing a diverse crop mix (\( i = d \)).

I assume that growing the monoculture goes hand in hand with pesticide use, which eliminates all pests from the field and achieves the highest possible yield per field \( y_{\text{max}} \). This pesticide use costs price \( \pi \), expressed in units of yield for convenience. Growing the diverse crop mix goes hand in hand with lack of pesticide use: pest population in those fields is a function of landscape-level arthropod abundance, which increases linearly with the number of diverse crop fields across the grid. Under
maximal arthropod abundance (attained if all grid cells are planted with the diverse crop mix), I assume that the highest possible yield per field $y_{\text{max}}$ is achieved. Yield per crop field thus follows the following equations:

$$y_i = \begin{cases} y_{\text{max}} - \pi & \text{if } i = m \\ \frac{y_{\text{max}}}{2} \left(1 + \frac{x}{N}\right) & \text{if } i = d \end{cases}$$ \hspace{1cm} (3.1)$$

where $x$ is the number of fields in the grid that are planted with the diverse crop mix (including field $i$).

If the surrounding crop grid is fully planted with the monoculture, $x = 1$ and $y_{i=d} = \frac{y_{\text{max}}}{2} \left(1 + \frac{1}{N}\right)$. If the surrounding crop grid is fully planted with the diverse crop mix, $x = N$ and $y_{i=d} = y_{\text{max}}$. For now, I assume that $\frac{y_{\text{max}}}{2} \left(1 + \frac{1}{N}\right) < y_{\text{max}} - \pi < y_{\text{max}}$. In other words, an arthropod population supported by a single diverse field cannot compete with the efficiency of pesticides (net of pesticide cost), but an arthropod population supported by a fully diverse landscape exceeds the efficiency of pesticides (net of pesticide cost).

**Privately optimal choice for a single grid cell.** The farmer managing field $i$ will choose to plant the monoculture if $x < N \left(1 - \frac{2\pi}{y_{\text{max}}}\right)$, they will choose to plant the diverse crop mix if $x > N \left(1 - \frac{2\pi}{y_{\text{max}}}\right)$, and they will be indifferent between the two for $x = N \left(1 - \frac{2\pi}{y_{\text{max}}}\right)$. From the assumption of $\frac{y_{\text{max}}}{2} \left(1 + \frac{1}{N}\right) < y_{\text{max}} - \pi < y_{\text{max}}$, we know that $1 < N \left(1 - \frac{2\pi}{y_{\text{max}}}\right) < N$.

If each farmer chooses their crop, taking other farmers’ crops as fixed, this generates three equilibria: one stable equilibrium at $x = 0$ (low diversity equilibrium), one unstable equilibrium at $x = N \left(1 - \frac{2\pi}{y_{\text{max}}}\right)$, and one stable equilibrium at $x = N$ (high diversity equilibrium).

**Socially optimal choice for a single grid cell.** I assume now that the farmer managing field $i$ internalizes the arthropod externality when choosing which crop to plant. Specifically, they consider both their private yield, and the additional yield endowed to the other diverse fields if $i = d$, thanks to the increase in arthropod abundance. The comparison is now between $y_{\text{max}} - \pi$ and $\frac{y_{\text{max}}}{2} \left(1 + \frac{x}{N}\right) + (x - 1) \times \frac{y_{\text{max}}}{2N} = \frac{y_{\text{max}}}{2} \left(1 + \frac{2x - 1}{N}\right)$.

In this setup, if each farmer chooses their crop, taking other farmers’ crops as fixed, this generates again three equilibria: one stable equilibrium at $x = 0$ (low diversity equilibrium), one unstable equilibrium at $x = \frac{1}{2} N \left(1 - \frac{2\pi}{y_{\text{max}}}\right) + \frac{1}{2}$, and one stable equilibrium at $x = N$ (high diversity equilibrium).
We know that \[ 1 < N \left( 1 - \frac{2\pi}{y_{\text{max}}} \right) < N, \]
so the intermediate equilibrium is well defined \((x > 1)\),
and it occurs at a strictly lower level of diversity, compared to the private case where the externality
is not internalized.

**Socially optimal choice for all grid cell.** I now assume that a social planner chooses crop plantings
simultaneously across all grid cells. Total yield across the grid can be written as follows:

\[ Y(x) = (N - x)(y_{\text{max}} - \pi) + x \frac{y_{\text{max}}}{2} \left( 1 + \frac{x}{N} \right) \]  
(3.2)

where again \(x\) is the number of fields in the grid that are planted with the diverse crop mix.

\[ Y(x = 0) = N(y_{\text{max}} - \pi), \quad Y(x = N) = Ny_{\text{max}}, \quad \text{and} \quad Y'(x) = \pi - \frac{y_{\text{max}}}{2} + y_{\text{max}} \frac{x}{N} > 0 \text{ for all } x, \]

which implies that a fully diverse grid \((x = N)\) is socially optimal in this setup.

What this model has shown is not so surprising. If local pesticide benefit (net of pesticide cost)
exceeds local arthropod benefit, but is lower than landscape-level arthropod benefit, then full reliance
on arthropods via complete planting of a diverse crop mix is socially optimal. If individual farmers
choose their crop separately (and non-simultaneously), there can be multiple equilibria: two stable
equilbria at both the low and high diversity extreme, and one unstable equilibrium at an interme-
diate level of diversity. Internalizing the arthropod externality shifts the unstable equilibrium to a
lower level of diversity. This doesn’t matter much in practice since this equilibrium is likely never real-
ized, but it means that the high diversity equilibrium will arise from a larger set of initial grid patterns.

**Extensions.** I have assumed that yield increases linearly with arthropod abundance. I relax this
assumption and assume instead decreasing returns to arthropods. This does not affect the results,
as long as the supremacy of the landscape-level arthropods over pesticides is maintained. The only
change is that the unstable intermediate equilibrium now occurs at a lower or higher level of diversity,
depending on the position of the concave arthropod abundance curve (as a function of diverse fields
\(x\)), relative to the original linear curve.

I have assumed that both the monoculture and the diverse crop mix share the same maximum
yield level \(y_{\text{max}}\). I relax this assumption and assume that there is a comparative advantage to the
monoculture in this landscape. Essentially, \(y_i = y'_{\text{max}} - \pi \) if \(i = m\), where \(y'_{\text{max}} > y_{\text{max}}\). It could
still be the case that diversity beats monoculture, if the productivity loss from switching to the
diverse mix is small or pesticide costs are large. If so, then the dynamics would be unchanged (aside from the position of the unstable intermediate equilibrium). However, it could also be the case that monoculture becomes a superior strategy, even relative to landscape-level arthropod abundance. In that case, monoculture would be the privately and socially optimal choice, and a full monoculture grid would be the unique equilibrium.

Incorporating the social costs of pesticide use (e.g. from water pollution), which could be internalized by individual farmers or by the social planner, would shift the intermediate equilibrium to a lower level.

Incorporating economies of scale within a field is similar to assigning a comparative advantage to the monoculture. Incorporating economies of scale across fields would require to assume that farmers own multiple fields across the grid or share resources with their neighbors. I expect that this won’t change much to the overall dynamics, except that the private equilibrium might be closer to the socially optimal one. Whether the socially optimal grid is low diversity of high diversity depends on the strength of economies of scale.

I also allow arthropod and pesticide effects to interact. I first assume that the presence of arthropods decreases the amount of pesticide required on monoculture fields, and then assume that pesticides have a spatial externality and decrease the population of arthropods across the grid. In both cases, monoculture fields become relatively more competitive and the intermediate equilibrium shifts to a higher level of diversity \( x \). If landscape-level arthropods remain more efficient than pesticides net of their cost, the social optimum is unchanged.

It would be interesting to expand the choice set for the farmer, and consider crop diversity across the entire crop grid, not just at the level of a field. This is my next step.

This model is probably too simplistic. It leads to clear cut outcomes, and does not enlighten much the question of how choices to incorporate crop diversity benefits impact the privately and socially optima for crop diversity. It does illustrate how a coordination failure can arise in this setup, across neighboring farms: farmers can get stuck in a low diversity equilibrium, even if the high diversity equilibrium is socially optimal. This creates path dependence, as historical crop diversity choices dictate which of the low-diversity and high-diversity equilibrium gets realized.
3.5 Conclusion

This chapter has tried to sift through the rich literature on crop diversity impacts on yield, and the more limited literature on optimal crop diversity, to assess whether perceived economic benefits from crop diversity could be disconnected from potential benefits from crop diversity. Some specificities of crop diversity benefits are addressed to some extent in the literature on optimal crop diversity, such as spatial externalities from natural pest control, or the risk of widespread crop failure from a pathogen attack. There does seem to remain a gap: functional forms are often disconnected from the ecology literature and therefore do not allow to dig much into how crop diversity processes impact crop yield, spatial and temporal scales do not always match the crop diversity channel depicted. This gap could lead to the underestimation of socially optimal levels of crop diversity, and the design of inappropriate agricultural policies like crop insurance programs, or trade regulations.

Both the misperception of crop diversity benefits and path dependence in crop diversity choice could explain why crop diversity seems to be under-represented in some of the largest crop production systems. Other reasons are likely at play: farmers in the Corn Belt mention in particular path dependence due to economies of scale or sunk investment in tools and skills, and lack of markets for non-dominant crops (Roesch-McNally et al., 2018).
Bibliography


Elton, C. et al. (1958). The ecology of invasions by animals and plants. *The ecology of invasions by animals and plants*.


