Fusing satellite data to monitor the urban area's effect on plant phenology

Norman Gervais
University at Albany, State University of New York, inf.gervais@gmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Environmental Sciences Commons, and the Remote Sensing Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/2060

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive.
Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
Fusing Satellite Data to Monitor the Urban Area’s Effect on Plant Phenology

by

Norman Gervais

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Emergency Preparedness, Homeland Security, and Cybersecurity
2018
Preface

The contents of Chapter 2 have been previously published in its entirety under a Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/) and were then reformatted and restyled, including but not limited to renumbering tables and figures, for this dissertation. This previously published article is included in this dissertation because the two chapters that follow it (Chapters 3 and 4) build on the research outlined in it. As indicated by the citation below, I was the lead researcher for the article. The citation to the published article:

Acknowledgements

I am thankful for all of the help and guidance that my committee, Dr. Alex Buyantuev (chair), Dr. John Davis, and Dr. James Mower, provided during the dissertation process. I would also like to thank all of my family.
Abstract

This dissertation centrally focuses on developing and validating the use of fused satellite imagery to monitor the effects of the urban area on plant phenology, specifically the timing of the start and end of the growing season (SOS and EOS, respectively). In the first paper, Chapter 2, data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat were fused together using the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) to produce a time series of a high spatial and high temporal resolution vegetation index. From this time series, the SOS and EOS were extracted and compared between the urban and exurban developed areas. This paper used two versions of the fusion software and the results from both versions generally supported the hypothesis that the urban areas have an earlier SOS, later EOS, and therefore a longer length of season (LOS). In addition, the 30m Landsat-like resolution of the fused imagery allowed differentiation between urban and exurban areas at a much finer spatial resolution, allowing for a better understanding of heterogeneous effects than would be possible at a coarser resolution of MODIS (500m). However, two issues were identified. First, the two software versions showed minor differences in results. Without in-depth validation of the software, it was not known which version produced the most reliable results and therefore which version should be adopted in future research. Second, the paper did not attempt to compare the results with ground conditions. Therefore there was no evidence to support that the observed parameters were actually related to true SOS or EOS, or if they were related to some other phenomena. The second and third papers investigated these two issues. The second paper focused on a technical comparison of the ability of the original and the most recent version (v1.2.2) of the STARFM software to predict high temporal and high spatial resolution fused images. The results show that v1.2.2 can be adopted. The third paper used v1.2.2 to produce
fused images for three US cities that coincided with in situ citizen science phenology data of the red maple (*Acer rubrum*). Comparison of the phenophase dates derived from the fused imagery with these in situ data showed that the amount of relationship varied between the two depending on the phenophase and city. In addition, although usually less than half of the variation in the phenology produced from the fused data was explained by the observed phenology of the red maple, there were statistically significant relationships between the two.
Table of Contents

Chapter 1: A Review of Plant Phenology: Its Importance, Drivers, and Observations Across Urban Areas

1. Introduction
2. Importance and Causes of Phenology
   2.1. Importance of Phenology
   2.2. Biophysical Drivers of Phenology
3. Cues for Phenology
   3.1. Growth Cessation
   3.2. Budburst
4. Observing Phenology
   4.1. Citizen Science Observations of Phenology
   4.2. Satellite Observations
   4.3. Limitations of Satellite Imagery
5. Image Fusion
   5.1. STARFM Parameter Selection
6. Summary and Dissertation Outline
7. References


1. Introduction
2. Materials and Methods
   2.1. Study Area
   2.2. STARFM Model Description
   2.3. Satellite Data Processing
   2.4. STARFM Parameters and Input Text Creation
   2.5. NDVI Time Series Preparation and Seasonality Extraction
   2.6. Dates of Phenological Events for Developed Areas
   2.7. Statistical Analyses
3. Results
   3.1. Accuracy Assessment of STARFM
   3.2. Descriptive Statistics
   3.3. Tests for Variance Equality and Normality
   3.4. Tests for Inequality of Phenological Parameters
4. Discussion
   4.1. Start of Season
4.2. End of Season
4.3. Length of Season
4.4. Study Assumptions and Potential Limitations
4.5. Future Work
5. Conclusions
6. Supplementary Materials
7. Acknowledgements
8. Author Contributions
9. Conflicts of Interest
10. References

Chapter 3, Paper 2: Examining the upgraded STARFM: Quality and Quantity of Predictions

1. Introduction
2. Methods
   2.1. Image Fusion
   2.2. Validation
3. Results
4. Discussion
   4.1. Quality and Quantity of STARFM Predictions
   4.2. Limitations
5. Conclusions
6. References

Chapter 4, Paper 3: Validating Urban Phenology from Fused Time Series Imagery with Citizen Science Data for *Acer rubrum*

1. Introduction
2. Methods
   2.1. Study Area
   2.2. In Situ Data
   2.3. City and Species Selection
   2.4. USA-NPN Phenophase Selection and Quality Control
   2.5. Image Fusion
   2.6. Extraction of Phenophase Events Dates
   2.7. Comparing Phenophases
3. Results
   3.1. Image Fusion Validation
   3.2. Variability of Phenophases
   3.3. Average Advancement of Seasons
4. Discussion
   4.1. Accuracy Assessment of STARFM
   4.2. Variability
   4.3. Average Normalized Values
   4.4. Limitations and Future Work
5. Conclusion
6. References

Chapter 5: Conclusions

1. General Conclusion
2. Existing Gaps in Knowledge and Future Work
3. References
Chapter 1
A Review of Plant Phenology: Its Importance, Drivers, and Observations Across Urban Areas

1. Introduction

The world’s urban population has been growing quickly since 1950. At that time, only 30 percent of the world’s population lived in an urban area. In 2014, 54 percent of the world’s population, or 3.9 billion people, lived in urban areas. This percentage is projected to keep increasing toward 66 percent by 2050 (United Nations, 2014), which makes it important to understand how the urban area interacts with the environment. One of the interactions is with the temperature, elevating it in the urban area, known in the literature as the urban heat island (UHI).

UHI is a concept that is used to describe a higher temperature in the urban environment compared to the surrounding rural areas (Oke, 1982). It is mainly a result of changing permeable landscapes into a man-made urban texture (Luo & Asproudi, 2015) and the effects of it arise mainly due to the modified surface affecting the storage and transfer of heat, water, and airflow (Hartmann et al., 2013). Across all seasons, there is a linear relationship between percent imperviousness and land surface temperature (LST) (Yuan & Bauer, 2007). The effects are measurable, with the annual mean air temperature in cities with at least 1 million people being 1-3 °C above their surroundings (US EPA, 2016). Even more extreme, Tokyo has been shown to be 8.1 °C higher than its suburb (Saitoh, Shimada, & Hoshi, 1996). However, even cities as small as 2-3 km² have shown an UHI effect of 0.17 °C, with a larger effect of 2.9 °C in the day and 2.3 °C in the night in larger cities of >100 km² (M. Tan & Li, 2015). Because the UHI effect is comparable to projected increases in global temperature (Jochner & Menzel, 2015), urban
environments have been used as a surrogate for impacts of future climate change (i.e. Ziska et al., 2003). Based on an array of data, biologists suggest that climate changes of varying magnitudes can have big effects on species (Wuethrich, 2000).

Phenology is described as “the timing of seasonal activities of animals and plants” (Walther et al., 2002, p. 389) or the “science of recurring events in nature” (Badeck et al., 2004, p. 295). It is a process that can be used to track responses to climate change (Walther et al., 2002). For example, there is a documented relationship between urban intensity and land surface phenology (LSP) (Walker, de Beurs, & Henebry, 2015). LSP is the timing of changes in electromagnetic radiation reflectance from the land surface, caused by foliage change, measured with remote sensors (Hanes, Liang, & Morisette, 2014). Studies using remote sensing have shown an extended growing season in urban areas (Jochner & Menzel, 2015). In addition, it was shown that not only is the urban core influenced by the UHI, but also the surrounding region. However, the influence on phenology decays exponentially with distance (Zhang, Friedl, Schaaf, Strahler, & Schneider, 2004). During a warmer spring leaves will appear earlier (Wesolowski & Rowiński, 2006) and plants will usually develop earlier in urban areas by a factor of days to weeks (Jochner & Menzel, 2015). However the difference in the timing of the end of season (EOS) between urban and rural areas is highly variable (White, Nemani, Thornton, & Running, 2002) and is not always distinct (Jochner & Menzel, 2015). In the fall, growth cessation and dormancy are mainly in response to the shortening of the photoperiod (Perry, 1971), defined as the length of sunlight exposure time per day (Cleland, Chuine, Menzel, Mooney, & Schwartz, 2007), although growth cessation can also be caused by low temperatures (Heide, 1974). In addition to variation in the temporal gradients in leaf onset (Fisher, Mustard, & Vadeboncoeur, 2006), variation in intra-urban phenology has been found (Zipper et al., 2016), with the amount
of impervious surface area around vegetation patches influencing start of season (SOS) timing (Melaas, Wang, Miller, & Friedl, 2016).

Comparing urban and rural trends of phenology allows for phenological trends to be evaluated in different temperature conditions and assessment of the phenological impacts of climate change (Jochner & Menzel, 2015). Satellite measurements have been used to study phenology over large areas (Hanes et al., 2014), however, due to clouds, atmospheric conditions, and infrequent satellite coverage this type of analysis has traditionally been restricted (Holben, 1986). Various methods have been developed to screen for cloud-contaminated pixels (Nixon, Wiegand, Richardson, & Johnson, 1982). Also, there is a relatively long history of atmospheric correction for Landsat Thematic Mapper (TM) imagery (S. Liang, Fang, & Chen, 2001). In addition, with the goal of obtaining more information, by combining different data, than can be obtained from each data by itself (Pohl & Van Genderen, 1998a), blending algorithms have been developed (Emelyanova, McVicar, Van Niel, Li, & van Dijk, 2013). One method, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM), was designed to predict surface reflectance values with Landsat spatial resolution and Moderate Resolution Imaging Spectroradiometer (MODIS) temporal resolution (Gao, Masek, Schwaller, & Hall, 2006). Because the urban area consists of small inter-city features and is thus heterogeneous, STARFM has the potential to develop a better model of observing urban phenology.

2. Importance and Causes of Phenology

Phenological research, from a biological perspective, mainly addresses the timing of the changes in recurring phases of organisms. In addition, some phenological research addresses continuous changes, such as the continuous change in remotely derived reflectance values of
vegetation (Badeck et al., 2004). Although Jochner and Menzel (2015) exemplify that there are a wide variety of factors which influence phenology, the remainder of the current paper focuses on the role of temperature as the primary cue of plant phenology. Temperature explains 70% of the variation in bud break (Menzel & Fabian, 1999). By studying the current relationship between phenology and temperature, we may be able to make better predictions of future changes in phenology as a result of this changing climatic variable.

Plant phenology varies with the annual course of weather (Badeck et al., 2004). For example, in cool and temperate regions, trees have adapted to annual weather variations. Since plants cannot be active and cold resistant simultaneously (Hänninen & Hari, 1988), trees need to extend the growth period to maximize growth while still avoiding frost damage. They do this through regulating the timing of bud burst in the spring and growth cessation and dormancy in the late summer and fall (Häkkinen, Linkosalo, & Hari, 1998).

For deciduous trees, a new growth of leaves is needed to initiate photosynthesis. Then in autumn, to recover mineral nutrients, the leaves undergo senescence before the first autumn frost (Lockhart, 1983). Plants become dormant in winter and are resistant to the cold, but the resistance is lost during the active period (Hänninen & Hari, 1988). These developmental stages, or phenological events, are referred to as phenophases (Cleland et al., 2007).

2.1. Importance of Phenology

Earlier green land cover and later autumnal events may alter biogeochemical processes and physical properties, leading to a change in seasonal climate (Peñuelas, Rutishauser, & Filella, 2009). Depending on the ecosystem type, phenological impacts on albedo will have varying impacts on climate change feedbacks (Richardson et al., 2013), contributing to solar heating of
the land (Bonan, 2008). In addition, working against temperature extremes, early green-up seasons reduce the rise of maximum surface temperatures and late green-up seasons have the opposite effect. This is expected because of the increase in energy usage from evapotranspiration and also the increase in heat capacity from the added water vapor (Schwartz & Karl, 1990). These factors can affect temperature in already warmer cities. By better understanding the dynamics of plant phenology and temperature, there is the potential to plan for a more comfortable living environment.

In addition to the effects of phenophases on temperature, the length of the growing season also plays an important role in the amount of CO₂ in the atmosphere (Peñuelas & Filella, 2001). Gross primary production (GPP) and ecosystem respiration are major processes that control land-atmosphere CO₂ exchange, and terrestrial GPP is the largest global carbon flux (Beer et al., 2010). Earlier spring leads to an increase in GPP and ecosystem respiration (RE), with the increase in GPP being larger than the increase in RE (Richardson et al., 2009). In addition, the net ecosystem CO₂ exchange (NEE) increases with each day that the growing season lengthens (Baldocchi et al., 2001). Therefore, it is important to understand how the expanding urban environment will impact climate change through altering plant CO₂ uptake.

Beyond abiotic effects, phenophase timing also has direct effects on organismal factors of the ecosystem. For example, because the response to warming temperatures differs between species, phenological mismatches across trophic levels may already be occurring due to a different rate of shifting of co-occurring organisms, changing the biological communities (Primack et al., 2009). This could potentially lead to extinctions.
2.2. Biophysical Drivers of Phenology

For new leaves, there is a potential for frost injury, but there is also an assumption that earlier growth increases potential photosynthetic yield (Lockhart, 1983). It has been shown that the amount of dry matter which a plant stand produces is linearly related to the amount of light energy intercepted by the foliage canopy. There is also a near linear decrease in light interception with later canopy development, which makes early canopy development important (Cannell, 1989). These two parameters of potential for frost injury and increased photosynthetic yields, acting in opposite directions, normally determine the best time for shoot growth. Therefore, the timing of growth initiation of new leaves should be based on a balance between the increased risk of frost injury and annual productivity (Lockhart, 1983).

However, Lockhart (1983) proposed that the optimal annual initiation time of bud growth is not only dependent on climate, but also the genetic character of frost resistance. If two strains that respond to temperature were transferred to the same location from different areas, the predicted behavior of transferred strains from higher latitudes would flush earlier at the same location than strains that were transferred from a warmer climate (Lockhart, 1983).

The timing of growth cessation for species with intermediate growth is a necessary climatic adaptation. It is assumed that trees maximize the use of growth resources if they have relatively late cessation, but they may also be damaged or killed by frost (H. Hänninen, Häkkinen, Hari, & Koski, 1990). Although evergreens retain foliage throughout the year and therefore have the potential for continuous carbon assimilation (Schaberg, Wilkinson, Shane, Donnelly, & Cali, 1995), field photosynthesis closely parallels seasonal temperatures (Schaberg, Shane, Cali, Donnelly, & Strimbeck, 1998) where winter photosynthetic rates have been shown to be generally low, although they can increase for prolonged thaws (Schaberg et al., 1995).
To maximize gross carbon assimilation, leaves would continue to photosynthesize until killed by the first autumn frost. However, a well-defined senescence process starts before the first expected frost and may start earlier in areas of scarce mineral nutrients (Lockhart, 1983). In addition, in northerly latitudes total interception is insensitive to leaf fall timing since at that time the incoming solar radiation is small (Cannell, 1989). In the autumn, vegetative and floral buds enter a state of dormancy (Cannell & Smith, 1986). During the winter dormancy period, plants resist low temperatures and then when they reach the active period, their resistance to the cold is lost. Plants cannot be active and cold resistant simultaneously, therefore they need to find a compromise between the risk for injury and the length of activity (Hänninen & Hari, 1988).

In summary, in environments where there is a risk for damage due to freezing temperatures, plants aim to maximize growth but cannot grow year around due to the potential for injury caused by the cold. Instead, they need to maximize their growing season to outcompete other species while avoiding the potential of damage by freezing temperatures. To do this, plants need to regulate not only the start of their growing season but also the end of it. In the spring, the plants need to start growth at the first chance they can while minimizing potential damage from late frosts. Then in the fall, they need to delay the cessation of growth as late as they can while still allowing time to prepare for the winter season.

3. Cues for Phenology

Since plant phenology varies with the annual course of weather, one might expect it to be one of the most responsive aspects of nature to the climate (Badeck et al., 2004) and to occur by a variety of natural and artificial agents (Perry, 1971). Spring warming, winter chilling, and photoperiod all can influence budburst of temperate trees (Hunter & Lechowicz, 1992), with
various factors being interdependent (Campbell & Sugano, 1975). At the end of the growing season, dormancy and growth cessation are responses to photoperiod, however other factors such as temperature may bypass the control of photoperiod (Perry, 1971). In addition, plants themselves also regulate their own functions. Provided there are no major differences in temperature conditions, budburst occurs synchronously if the genetic properties of plants are similar (Hänninen & Hari, 1988). This fact that individual species regulate their own functions demonstrates the importance of studying responses to cues at the species level, as many field studies do. However, for temperate regions the responses and cues appear to be similar (Table 1-1) which allows us to generalize our expected response of landscape level phenology to the UHI across temperate regions.

<table>
<thead>
<tr>
<th>Action</th>
<th>Cue</th>
<th>Relationship</th>
<th>Example of study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Budburst</strong></td>
<td>Temperature</td>
<td>Budburst comes after a set temperature accumulation, earlier budburst with warmer springs</td>
<td>Hunter &amp; Lechowicz, 1992 and Wesołowski &amp; Rowiński, 2006</td>
</tr>
<tr>
<td></td>
<td>Winter Chilling</td>
<td>Increased winter chilling decreases time to budburst, to some minimum value</td>
<td>Cannell &amp; Smith, 1986</td>
</tr>
<tr>
<td></td>
<td>Photoperiod</td>
<td>Increase in photoperiod leads to earlier budburst. Potentially species specific.</td>
<td>Partanen, Koski, &amp; Hänninen, 1998 &amp; Basler &amp; Körner, 2012</td>
</tr>
<tr>
<td><strong>Growth Cessation and Dormancy</strong></td>
<td>Temperature</td>
<td>Low temperature causes growth cessation</td>
<td>Heide, 1974</td>
</tr>
<tr>
<td></td>
<td>Photoperiod</td>
<td>Short days causes growth cessation</td>
<td>Heide, 1974</td>
</tr>
<tr>
<td></td>
<td>Temperature and Photoperiod</td>
<td>High temperatures accelerate response to short days</td>
<td>Heide, 1974</td>
</tr>
</tbody>
</table>

Table 1-1. Summary of frequently identified cues and their relationships to either budburst or growth cessation and dormancy in temperate regions.
3.1. Growth Cessation

In autumn the leaves undergo a senescence process, which starts before the first autumn frost (Lockhart, 1983), and vegetative and floral buds enter a state of dormancy (Cannell & Smith, 1986). Although some avoid a restrictive definition of dormancy, also referred to as rest, a dormant plant has a period of noticeable reduced growth rate with little or no cell division in terminal and lateral meristems along with a winter chilling requirement (Perry, 1971).

During the first stages of dormancy, the active growth and formation of new leaves stops before any danger of frost, cellulose formation is slowed, lignin formation is accelerated, fats and starches accumulate in the storage tissue, buds grow and take winter form, and the earliest stage of leaf development for the next growing season are formed. The leaves remain green and can still photosynthesize, but cannot resume growth at this point. Then, as night temperature fall, enzymes start to digest cell walls of areas that will be lost, the chlorophyll starts to break down, and some minerals move to the permanent organs. The leaves then fall and the tree can then withstand winter conditions (Perry, 1971).

There is an assumption that the shortening of photoperiod is the main driver of growth cessation and dormancy (Perry, 1971). However, it has been shown that growth cessation can be also caused by low temperature, and high temperature accelerates the response to short days (Heide, 1974). In addition, it has been suggested that light intensity, nutrient concentration, soil moisture, and shock treatments may also override photoperiodic control (Perry, 1971). Therefore, the effect of the UHI on fall cessation is not entirely clear. However, based on these assumptions, one may reasonably expect the UHI either to have no effect on cessation, because UHI has no influence on photoperiod, or potentially to accelerate the plants’ response to shortening days.
Trees then will remain in this state of rest until they have been exposed to a period of cold weather, referred to as a chilling requirement. The amount of chilling required varies with species and weather of the previous season (Perry, 1971). The dormancy is decreased by chilling temperatures, and with increased chilling there is a decrease in thermal time to budburst (Cannell & Smith, 1986). However, not all species have a true chilling requirement (Perry, 1971). Assuming the UHI does not raise temperatures above the chilling requirement, this should have no effect on budburst.

3.2. Budburst

The timing of spring events is mainly regulated by temperature (Linderholm, 2006). Although not all responses of biological processes to temperature are identical, it is still important to develop general theory to understand and describe observed phenomena (Johnson & Thornley, 1985). For example, the timing of phenological events such as bud burst may be species-specific, however some species have been shown to respond to environmental variation by shifting their onset dates in parallel, conserving the relative order of bud burst. However, at the individual level, it has been shown that within a species individual plants may have their own environmental requirements. Even though there may be individual variation in how the cues are interpreted, a large body of data have shown that bud burst and leaf growth rate are controlled strongly by temperature, where leaves appear earlier and have faster growth during a warmer spring (Wesołowski & Rowiński, 2006). Therefore, we could expect to see the timing of budburst for any given species to shift to an earlier period in response to UHI-induced temperature increases.
Using the centuries old heat unit approach, relationships between plants and temperature are studied by the accumulation of mean daily temperatures above some threshold. The difference between a base temperature and the actual temperature is termed a degree day or heat unit. Heat units accumulate over time (also referred to as heat sums, growing degree-days, etc.) until a required threshold is reached (Wang, 1960), at which point it is assumed that an event occurs (Hunter & Lechowicz, 1992). Variation in the timing of phenological events between years is large, but has been explained by variations in heat sums for many phenological phases. Therefore, a relationship between the onset of a phenophase and recent heat sums is expected (Badeck et al., 2004). However, it is only after a winter chilling requirement is met that the trees will start growth in response to a heat sum threshold that is above the chilling temperature (Perry, 1971).

Even though the best predictor of bud burst has been shown to be heat sums (Wesołowski & Rowiński, 2006), the effect on bud burst from temperature is highly interdependent with chilling and photoperiod, both of which were shown to widen the response to temperature by increasing the rate of bud burst (Campbell & Sugano, 1975). For example, some species appear to display sensitivity to photoperiod by delaying budburst with shorter photoperiods (Basler & Körner, 2012) and speeding it up with an increased photoperiod (Partanen et al., 1998). Yet other species are insensitive to photoperiod (Basler & Körner, 2012) or it has negligible effects on budburst date (Cannell & Smith, 1983).

4. Observing Phenology

Evidence for expected phenological trends come from both ground observations of individual plants and remotely sensed imagery (Badeck et al., 2004), as physical and
physiological parameters of vegetation can be obtained from satellites (Myneni, Hall, Sellers, & Marshak, 1995). However, ground observations and satellite-derived measures provide complementary information and measure qualitatively different traits (Badeck et al., 2004) at different scales.

Traditionally, phenological events have been observed using in situ observations. Data derived via this method are useful, but are not spatially and temporally extensive. This inhibits systematic assessment of larger areas (Hanes et al., 2014). Ground observations of phenophases are conducted at the level of individual plants. They generally do not represent a mixture of species in a landscape. The date of an event (i.e. date of bud burst), is determined by observing individual plants in a well-known microenvironment. These observations can be scaled to the species level only if biological variability and microclimate influences are assessed. In contrast, remote sensing provides area-averaged information based on the reflectance (Badeck et al., 2004).

### 4.1. Citizen Science Observations of Phenology

One option to collect fine-grained information over large areas and long times is to use citizen science (McKinley et al., 2015). Citizen science is referred to as “the practice of engaging the public in a scientific project – a project that produces reliable data and information usable by scientists, decisionmakers, or the public and that is open to the same system of peer review that applies to conventional science” (McKinley et al., 2015, p. 3). The investment into citizen science is generally done not only to engage the public, but also to do science that may not otherwise be feasible due to scale or other reasons (McKinley et al., 2015).
Participation by the general public will be a necessary component of successful phenological monitoring activities. In 2007, the contemporary USA National Phenology Network (USA-NPN) of individual and organizational partnerships was established to collect common phenology data at the national level. Through this network the National Coordinating Office (NCO) acts as a coordination center and maintains data and information sharing needs of the network. To enable collection and organization of phenology observations by individuals and groups, the NCO developed a web application called Nature’s Notebook (NN), which provides standardized protocols for phenology monitoring (Schwartz, Betancourt, & Weltzin, 2012). The protocols quantify onset, duration, and intensity of phenophases of both plants and animals by recording the status of the presence or absence of a phenophase which can then be used to determine the date of an event (Denny et al., 2014). The NCO also makes these data freely available online (Schwartz et al., 2012).

Using protocols similar to the USA-NPN program, Fuccillo et al. (2015), compared phenology data collected by volunteers to that collected by a trained plant ecologist using the same protocols for the same individual plants. It was shown that volunteers correctly identified phenophases 91.3% of the time, with unfolded leaves being accurately identified 95.6% correct compared to 81.3% for emerging leaves, open flowers 87.6% compared to 85% for full flowers, and 99.3% for ripe fruits. However, the average accuracy of phenophase transition dropped to 70.2%. Since this study, protocols have been refined. They concluded that phenology monitoring programs, such as Nature’s Notebook, can result in high-quality data (Fuccillo, Crimmins, de Rivera, & Elder, 2015). In addition, quality assurance measures have been implemented to minimize inaccurate entries into the database of the Nature’s Notebook (Rosemartin et al., 2014).
Using the observations for four *Quercus* species that were contributed to the USA-NPN during 2009-2014, Gerst, Rossington, & Mazer (2017) observed the difference in responsiveness to climatic and geographic drivers. Data used were limited to “breaking leaf buds” and “flowers or flower buds.” To help ensure the accuracy of seven days of maximum error, data were filtered to include only onset day of year (DOY) that had a positive observation recorded within one week after a negative observation. In addition, to limit analysis to spring phenology only DOY observations between 1 and 182 were used (Gerst, Rossington, & Mazer, 2017), similar to the 180 DOY threshold used by Piao et al. (2015), or leaf observations beyond DOY 150 and bloom observations beyond DOY 180 (L. Liang & Schwartz, 2014). This is also in line with the spring limitation of DOY 50-200 and autumn limitation of DOY 200-365 (Elmore, Stylinski, & Pradhan, 2016), and the exclusion of data for plants that were reported to flush later than the end of June (Y. H. Fu et al., 2015). Last, if multiple individuals were monitored at the same location, the site mean was used. By doing this, they were able to see species specific responses to climate drivers within and between ecoregions, as well as the influence that geographic variables had on phenophases (Gerst et al., 2017). In separate studies since 2007, various applications of the Spring Indices (Schwartz, 1997) have been implemented (i.e. (McCabe, Ault, Cook, Betancourt, & Schwartz, 2012) and the USA-NPN data were used to:

- Validate a model which predicted leaf emergence (Medvigy, Kim, Kim, & Kafatos, 2016)
- Evaluate model performance of continental scale budburst and dormancy (Yue, Unger, Keenan, Zhang, & Vogel, 2015)
- Develop models aimed at examining the association of climatic variables and onset dates via stepwise regression (Mazer, Gerst, Matthews, & Evenden, 2015)
- Validate phenological predictions in the development of a distribution model (Chapman, Haynes, Beal, Essl, & Bullock, 2014)
- Calibrate a leaf coloring model (Jeong & Medvigy, 2014)
- Compare root phenology to leaf phenology (McCormack, Gaines, Pastore, & Eissenstat, 2015)
- Estimate heat requirement for flushing (Y. H. Fu et al., 2015)
• Help parameterize a dynamic vegetation model by plant functional types and compare to a model that uses the same leaf-out for all species (Euskirchen, Carman, & McGuire, 2014)
• Assess the direct effect of climate on phenology from all other effects, including genotype (L. Liang & Schwartz, 2014)
• Develop and validate budburst models that were used in predicting future spring phenology changes (Jeong, Medvigy, Shevliakova, & Malyshev, 2013)
• Compare nighttime and daytime temperature influence on leaf unfolding dates based on the in situ data to satellite NDVI predictions (Piao et al., 2015)

The USA-NPN has also been used to evaluate dates of MODIS-derived spring green-up dates. Peng et al. (2017) found satellite-obtained vegetation indices (see next section) are generally consistent with the USA-NPN data for several land cover types, with the Enhanced Vegetation Index generally having a more significant correlation (Peng et al., 2017). The USA-NPN lilac (Syringa vulgaris) data have also been used to validate satellite derived Leaf Area Index (LAI) phenology (Verger, Filella, Baret, & Peñuelas, 2016). Elmore, Stylinski, & Pradhan (2016) addressed the degree of correspondence between the USA-NPN Nature’s Notebook (NN) data and MODIS for species observed at more than 30 sites and sites with five or more species observed. However, they relaxed this refinement criteria for both quaking aspen (Populus tremuloidies) and Red Rothomagensis lilac (Syringa × chinensis) since they are considered calibration species by the NPN. They also implemented two quality control (QC) steps. The first was limiting the date range of phenophases to an a priori species-specific threshold. The second was to remove all observations made by irregular volunteers by removing onset records not preceded by a ‘no’ within the previous ten days. A third QC component was developed to only include MODIS pixels that matched the NN data and had a forested pixel within the nine surrounding pixels, replacing the MODIS value with the median MODIS phenology date. The results of linear regressions (type 2) show significant correlation between NN and MODIS for poplar and lilac after the implementation of the first QC, improving with each additional QC
measure. In addition, they found that 105/228 of their models were significant (Elmore et al., 2016).

However, using more restrictive selection criteria comes at a cost of decreasing the sample size. When using three separate sets of criteria for data selection, ranging from using all “yes” observations regardless of a previous “no” to requiring a preceding “no” to most restrictive requiring a preceding “no” to be within seven days of the “yes”, it has been shown that the dataset is reduced by 27% and then by another 30%, respectively. The resulting sample sizes had little average effect on the differences and generally an insignificant effect for individual species, however the effect was significant when looking at all species together. When compared to latitude at the regional scale, the resulting estimated DOY across sample sizes did not statistically affect the slope (Gerst, Kellermann, Enquist, Rosemartin, & Den, 2016). After examination of a snapshot of the USA-NPN data, although it does have a wide temporal and spatial coverage, the sample of observations is still only a small fraction of the entire surface area. Therefore, the use of satellite data still has the benefit of being more spatially exhaustive.

4.2. Satellite Observations

To overcome this limitation of spatial and temporal extent, satellite derived measurements of the land surface reflectance have been used to study phenology over large areas (Hanes et al., 2014). The ability for satellites to observe surface radiation and monitor phenological cycles results in data being available for the entire globe for decades (Badeck et al., 2004), capturing long-term trends and interannual anomalies (Maignan, Bréon, Bacour, Demarty, & Poirson, 2008). Unlike individual in situ observations the remote sensing approach provides area-averaged information (Badeck et al., 2004). Ground point measurements are potentially not
representative of each pixel which may instead include a mosaic of vegetation (Zhang, Friedl, & Schaaf, 2006).

Satellite instruments measure the solar radiation reflected at certain wavelengths. For the remote sensing of vegetation, the red (0.6-0.7 µm) and near-infrared (0.75-1.35 µm) wavelength intervals have been the most helpful (Myneni et al., 1995), and even before 1980 it was common to apply combinations of the red and infrared wavelengths to monitor vegetation (Tucker, 1979). Green leaf has a chlorophyll absorption maximum at about 0.69 µm and lacks absorption at 0.85 µm. This leads to a strong contrast in absorption across the 0.65-0.85 µm wavelength interval. Vegetation indices (VIs) are able to capture this contrast by combining the red and near-infrared reflectance (Myneni et al., 1995).

The Normalized Difference Vegetation Index (NDVI) is the most widely used index (Myneni et al., 1995) among many other VIs (for example, see: Schultz et al., 2016), which individually have their own strengths and weaknesses (Huete, Liu, Batchily, & Leeuwen, 1997). NDVI is derived from images in the red and near-infrared (NIR) spectrum by using the following equation (Myneni et al., 1995):

\[
\text{NDVI} = (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED})
\]

This produces a range from -1 to 1, where negative values are related to areas of water and positive values are related to areas of green vegetation (Marchetti, Minotti, Ramonell, Schivo, & Kandus, 2016). NDVI allows us to estimate seasonal changes in phenological switches, cover from vegetation growth, and chlorophyll content (Badeck et al., 2004), as well as map vegetation traits and summarize the variability of plant phenology with the changing NDVI signal (Marchetti et al., 2016). The timing of the changes in reflectance is referred to as LSP (Hanes et al., 2014) and can be used to monitor differences between different time periods (de
Beurs & Henebry, 2004). When using this indicator for seasons, areas of sparse vegetation (i.e. NDVI <0.1) may be excluded (Piao et al., 2015).

There are various methodologies that use remote sensing to estimate phenology. In the evaluation of 10 methods to estimate the SOS, defined “as a rapid sustained increase in remotely sensed greenness after the longest annual period of photosynthetic senescence” (White et al., 2009, p. 2336), White et al. (2009) showed that when compared to ground data, although TIMESAT was outperformed, it still had a high $R^2$ with large bias toward early SOS (White et al., 2009). However, TIMESAT has several advantages, including being open source and having three smoothing functions (B. Tan et al., 2011). These advantages may offset the performance disadvantage.

TIMESAT is a computer program that smooths time-series satellite data to extract phenological parameters (Jönsson & Eklundh, 2002). Within TIMESAT, the dates of a phenological event are identified with a threshold method, with the default value of 20% of the seasonal amplitude (B. Tan et al., 2011). Of the three smoothing functions, the asymmetric Gaussian and double logistic functions use semi-local methods whereas the Savitzky-Golay method uses a local polynomial function to fit data. The Gaussian and logistical methods are less sensitive to noise, producing similar results with the Gaussian method being less sensitive to time-series data with gaps (Gao et al., 2008). It has also been shown that although a double logistic method and asymmetric Gaussian function also produced similar results for NDVI time series in the boreal and tundra, the double logistic function had a lower RMSE than the asymmetric Gaussian function (Beck, Atzberger, Høgda, Johansen, & Skidmore, 2006) and has since been used to derive green-up onset dates (i.e. L. Liang et al., 2014).
One quality criteria of TIMESAT is that the data cannot have too many missing data points. If this is the case, the time-series is discarded (Jönsson & Eklundh, 2002). Gao et al. (2008) proposed a method of gap filling to overcome this curve fitting failure by computing the gap pixel based on an appropriate seasonal-variation curve of high quality pixels of the same land-cover type that are nearby. If one is not found nearby, then an average of all high-quality seasonal-variation curves for the same land cover types is used. In addition, they applied weights in TIMESAT based on the MODIS QA layers, with higher weights for higher quality retrievals (Gao et al., 2008). Building on this work, Tan et al. (2011) explain that if the data gap which is causing the missing information is in the non-growing seasons (winter), the fit should still be made because the vegetation reflective characters should be relatively constant. In this case, the non-growing season was defined as a period of greater than a half of a month where the pixel is either snow-covered or the night surface temperature is equal to or less than 0 °C. They therefore replaced values in the non-growing season with a smoothed flat curve that connects the latest and first growing season points. The results attributed a low retrieval of the original TIMESAT in the northeast to long snow-covered periods where the data with missing values filled during the non-growing season succeeded with a higher retrieval ratio (B. Tan et al., 2011). Similarly, the MODIS snow/ice flag was used to identify all snow values and then replace them with the most recent snow-free value. Additionally, since the flag is not perfect, a screening of LST was used where any values of <5 °C were replaced with the nearest value. Additionally, single dates of missing data were replaced with an average of the two nearest neighbors with valid data and a three-point median-value moving-window was applied to all data (Zhang et al., 2006). Another solution, similar in nature, substituted a background index value for the index values that were contaminated with snow. The background value used was the minimum index value that was not
contaminated with either snow or clouds within the growing season or the maximum index value outside of the growing season, calculated from five good observations during the period where MODIS LST was <278 K. Short gaps caused by clouds were filled with a moving average value calculated from two neighboring good values and longer gaps of over a month were substituted with good observations from preceding or succeeding years (Zhang et al., 2017). Although these methods slightly vary, they all share one objective of filling in gaps of a time series to ensure TIMESAT can make a prediction.

In addition, distinguishing between the sources of variance in the signal is challenging, with an estimate for the date of an event being retrieved from the time course of the signal. For homogenous deciduous areas, critical periods usually can be clearly identified on the NDVI curve based on the maxima/minima. With heterogeneous areas, it may not be clear which event on the surface is changing the curve and an absolute threshold value which is related to phenological phases is easier to use (Badeck et al., 2004).

Validation of any remote-sensing analysis is always a concern. The most obvious and important way to assess phenology retrievals from a satellite is to validate against field observations, with large scale efforts such as the USA-NPN being able to possibly help toward collecting the data needed to fully assess quality (Zhang et al., 2006). Several factors may contribute to differences in the date of an event observed on the ground and the date of the event based on the time course of the signal. First, some models, which are used to account for spatial variation, and also the method and parameters used to identify an event, may introduce variance. Second, inter-individual variability of micro-climate, variation within the species itself, and observer error do not introduce systematic error in the determined event date. Third, when looking at bud burst of trees, heterogeneity will advance satellite measurements due to
understory plants, grasslands, agriculture, and gardening plants unfolding their leaves before
trees (Badeck et al., 2004).

4.3. Limitations of Satellite Imagery

Many studies that characterize urban morphology depend on satellite images (Van de Voorde, Jacquet, & Caners, 2011). However, the heterogeneity of the urban environment is a
problem for remote sensing (Herold, Scepan, & Clarke, 2002). The level of detail that maps
provide is dependent on the spatial resolution of the sensor, with a reduction in the problem of
mixed pixels with high spatial resolution sensors. Medium resolution images, such as Landsat or
SPOT, offer a lot of information for urban monitoring especially when historical perspectives are
required. However, this resolution is more suited for small-scale studies at the city-level rather
than intra-urban analysis. This is due to the fact that many intra-urban features are lost (Van de
Voorde et al., 2011) since we usually tend to think that the smallest object we can detect is
comparable in size with the instantaneous field-of-view (IFOV) of the instrument. However,
subpixel sized objects may still contribute to the received signal (Cracknell, 1998). This may be
important when observing urban phenology since it has been shown that growing season is not
uniform throughout the urban area and is longer near the urban core. However, the same study
also showed that larger contiguous areas in the urban area act similarly to the rest of the urban
area (Krehbiel, Zhang, & Henebry, 2017), providing evidence for the need of higher resolution
sensors to observe the smaller and fragmented areas in the urban environment.

Another issue is that the error in detecting phenology increases as the temporal resolution
is decreased (Zhang, Friedl, & Schaaf, 2009). Limited temporal resolution can constrain precise
predictions of phenology (Schwartz & Reed, 1999). Ideally, when some random noise is
introduced, aggregated time series reduce uncertainties and lead to minimum errors being produced when the temporal resolution is between 6-16 days (Zhang et al., 2009). However, even with sensors having this temporal resolution, due to clouds, atmospheric conditions, and infrequent satellite coverage remote sensing has traditionally been limited to few scenes over selected areas (Holben, 1986).

Various methods have been developed to screen for cloud contaminated pixels since their removal prior to interpretation would avoid erroneous conclusions about the conditions on the Earth’s surface (Nixon et al., 1982). Ultimately, however, these clouds, as well as sensor problems, can lead to missing values in time series data, and in a time series of 16-day data this greatly reduces the precision of detecting phenological transitions (Zhang et al., 2009). As a result, platforms with the medium resolution needed to observe the urban environment may not have a suitable number of annual images for observing phenology.

After removal of clouds, the additional constituents of the atmosphere may affect the signal by either scattering or absorption, which can either increase or decrease the value of the NDVI. Also, nearly transparent clouds may obscure the clear view of a surface (Holben, 1986). Therefore it is necessary to remove the atmospheric effects before applying satellite data to quantitative estimation (Jaelani, Matsushita, Yang, & Fukushima, 2015). The procedure of atmospheric correction retrieves what is called surface reflectance, which has a relatively long history consisting of several rough groups of methods (S. Liang et al., 2001).

For example, surface reflectance for Landsat TM and ETM+ are offered as Climate Data Records (CDR). This surface reflectance data is generated from the LEDAPS software (US Geological Survey, 2015b), which allows not only for multiple uses including mapping of land-cover and vegetation biophysics, but also comparison to reflectance data from other instruments.
such as MODIS (Masek et al., 2006). Surface reflectance generation has continued with the newest Landsat 8 via the LaSRC code (US Geological Survey, 2015a). These Surface Reflectance Data Products are available through the USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) On Demand Interface (US Geological Survey, 2015a).

Another method is used to process MODIS data. The MODIS Adaptive Processing System (MODAPS) generates MODIS surface reflectance data. This is then further refined to produce other products, such as a composited BRDF product (Justice et al., 2002), which can be found at several sources (NASA, n.d.). The BRDF, or Bidirectional Reflectance Distribution Function, describes the scattering of light from one direction into a different direction (Martonchik, Bruegge, & Strahler, 2000). Since Landsat imagery are acquired within 7.5 degrees off nadir, BRDF was not a concern when developing surface reflectance (Masek et al., 2006). In contrast to cloud removal, obtaining surface reflectance does not necessarily remove observations from a dataset, but instead normalizes the pixel values to a value as if they were recorded under the same environmental conditions. However, this does not overcome the spatial and temporal restrictions inherent to sensors.

5. Image Fusion

One way to overcome the limitation of data characteristics is to use data-data fusion or blending, which uses two data sources with complementary data frameworks and focuses on spatial and temporal dynamics (Emelyanova et al., 2013). The goal of image fusion is to obtain more information, by combining different data, than can be obtained from each data by itself (Pohl & Van Genderen, 1998b) and to produce imagery which better captures the spatio-
temporal dynamics by blending, for example, the high spatial resolution of Landsat with the high temporal resolution of MODIS. The results produce imagery with high spatial and temporal resolutions (Emelyanova et al., 2013).

However, traditional data fusion approaches may not be suitable for capturing changes in surface reflectance (Gao et al., 2006) or enhancing both the spatial and temporal resolutions at the same time (Zhu, Chen, Gao, Chen, & Masek, 2010). In 2006, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) approach was first published. It was designed to predict surface reflectance values with Landsat spatial resolution and MODIS temporal resolution (Gao et al., 2006), and is widely used as of 2013 (Emelyanova et al., 2013).

STARFM was originally shown, when using MODIS daily surface reflectance at 500m resolution and ETM+ data, to capture phenology changes over forested regions, gain temporal information from MODIS when looking at cropland, and capture patterns of complex mixed regions (Gao et al., 2006). The applications have since been applied to study, for example, the urban heat island via LST (Shen, Huang, Zhang, Wu, & Zeng, 2016), crop grown stages (Gao et al., 2017), NDVI values of semi-arid rangelands (Olexa & Lawrence, 2014), and community level phenology (L. Liang et al., 2014).

Since the introduction of STARFM, other fusion methods have built on this development to help improve on what it has accomplished. For example, the Spatial Temporal Adaptive Algorithm for Mapping Reflectance Change (STAARCH) outputs not only synthetic 30m surface reflectance predictions similar to STARFM, but also a date of disturbance for disturbed pixels (Hilker et al., 2009). However, these methods have difficulty when predicting reflectances of heterogeneous landscapes. To overcome this, the enhanced STARFM (ESTARFM) was developed (Zhu et al., 2010). However, this method has since been criticized as not taking into
account autocorrelation. As a result, mESTARFM was produced which modifies the selection process of similar pixels (D. Fu, Chen, Wang, Zhu, & Hilker, 2013). In addition, to incorporate sparse representation the Sparse-representation-based SpatioTemporal reflectance Fusion Model (SPSTFM) has been developed (Huang & Song, 2012). A bilateral filtering approach which incorporates all neighboring pixels (not just spectrally similar ones) has also been developed to model LST (Huang, Wang, Song, Fu, & Wong, 2013). Taking the strengths of STARFM and a Bayesian based unmixing algorithm into a separate fusion method, Spatial and Temporal Reflectance Unmixing Model (STRUM) was also created (Gevaert & García-Haro, 2015).

There are potentially even more new algorithms beyond these, however, the improvements are not clear. When originally validated, ESTARFM underperformed STARFM in predicting reflectance values for heterogeneous areas when STARFM used only one input base pair (Zhu et al., 2010). Therefore, it may be safer to stick to using the more simplistic STARFM as Schmidt et al. (2015) did. In addition, the acceptance of one base pair makes it suitable for study locations where high-quality Landsat images are limited (L. Liang et al., 2014).

5.1. STARFM Parameter Selection

Zhu et al. (2010) showed an improvement in STARFM’s ability to make predictions for heterogeneous areas when using only one base pair. Walker, de Beurs, Wynne, and Gao (2012) compared the results of using one base pair comprised of TM data and three different 500m MODIS datasets (daily reflectance data (MOD09GA), 8-day composite reflectance data (MOD09A1), and Nadir BRDF-adjusted (NBAR) reflectance 16-day composite data (MCD43A4)). They tested both forward and backward predictions by using two separate input pairs and observed several important results which can help to decide on base pairing. First, they
observed that out of each of the MODIS options for the course resolution data, the NBAR dataset
generally had the highest $R^2$ values and lowest absolute differences. Second, they were able to
explore the choice of MCD43A4 base pair date since they had two choices of MCD43A4 per
each Landsat input image. This is due to the fact that the MCD43A4 data product is taken over a
16-day period every 8 days, producing overlapping dates. Based on analysis of the different
options, they found that the choice of MCD43A4 input date only mattered for one of the input
dates. With only minor differences between the two MODIS dates when pairing with Landsat
data, the mean absolute difference was 2-3x greater when the earlier of the two overlapping
MCD43A4 images were used. They discussed that, in regards to phenological analysis,
STARFM “may be beneficial for bridging data collection gaps and providing the higher spatial
resolution necessary for analysis in dryland forest areas” (Walker et al., 2012, p. 392). In
addition, Walker et al. (2014) used TM and MCD43A4 data with one base pair to determine
STARFM’s ability to produce phenological metrics. They concluded that overall the STARFM
algorithm can make “valuable contributions to the study of dryland ecosystem phenology”
(Walker et al., 2014, p. 95). However, Gevaert & García-Haro (2015) have since shown that
STARFM captures NDVI values when using MCD43A4 and all available Landsat 8 data, but is
dependent on the number of input images and fails to capture phenological variations when using
one input image for an entire year. D. Fu et al. (2015) also found that selecting the nearest base
date to the prediction date will produce results with higher accuracy,

These studies can help determine the appropriate base pairs for STARFM. First, Walker
et al. (2012), make a clear case for using the MCD43A4 dataset. Second, based on Zhu et al.
(2010), Walker et al. (2012), and Walker et al. (2014) it is reasonable to use only one base pair
for each prediction. Third, Gevaert & García-Haro (2015) as well as D. Fu et al. (2015) have
made a strong case for using all available Landsat data in base pairing. Fourth, based on Walker et al. (2012), when choosing one of the overlapping MCD43A4 dataset to pair with Landsat, the results do not differ much, but do favor using the later dataset. Last, based on the reported individual values by Walker et al. (2012), the average $R^2$ values and the average MAD values were calculated and are better (by 0.0027 and 0.0005 respectively, unreported) when using the earlier prediction date.

6. Summary and Dissertation Outline

Since temperature is the main regulator for spring events (Linderholm, 2006) and also has a potential role in growth cessation (Heide, 1974), it is reasonable to expect that there would be a change in plant phenophase timing within the urban area due to the UHI, leading to longer growing seasons. This expectation has been shown to be true, with plants developing up to weeks earlier in cities (Jochner & Menzel, 2015). This is particularly important since plant phenology has a wide range of effects on the surrounding environment, ranging from changing the albedo (Richardson et al., 2013) and the surface maximum temperatures (Schwartz & Karl, 1990) to the amount of atmospheric CO$_2$ (Peñuelas & Filella, 2001). With the UHI effect being comparable to the projected increases in global temperature (Jochner & Menzel, 2015), the urban environment can be used as a surrogate for impacts of climate change (i.e. Ziska et al., 2003). Thus, understanding the effects of the urban environment on phenology is important not just in relation to that specific environment, but also in helping us understand the potential impacts of climate change across all systems. However, using a single satellite to observe this is limited due to the inherent trade-off that sensors have between increasing spatial resolution at a cost of decreasing temporal resolution.
In an attempt to overcome limitations in remotely sensed data, image fusion aims to combine two data sources with complementary frameworks (Emelyanova et al., 2013) and to obtain more information, by combing different data, than can be obtained from each data by itself (Pohl & Van Genderen, 1998b). This produces imagery with both high temporal and high spatial resolutions (Emelyanova et al., 2013). For example, STARFM is a fusion method that predicts surface reflectance values with Landsat spatial resolution and MODIS temporal resolution (Gao et al., 2006) and is widely used (Emelyanova et al., 2013). This method of image fusion applied to the observation of urban phenology should provide us with observations that have the spatial resolution appropriate to observe the urban environment and the temporal resolution to predict phenological events, leading to improved information on the potential impacts of future climate change. Chapters 2-4 built on this assumption.

Chapter 2, which has previously been published (Gervais, Buyantuev, & Gao, 2017), used STARFM to model the effects of the urban area on plant phenology. STARFM allowed the fusion of Landsat and MODIS time-series data, from which a vegetative index was calculated and inputted into TIMESAT for seasonality extraction. The main findings showed that this method generally produced results that would be expected, including an earlier urban start of the growing season, later urban end of growing season, and an overall longer urban growing season. However, this paper also found that the results varied depending on the base pair selection criteria and STARFM software version. It lacked a comparison to in situ data, so there was no way to determine which set or if any of the results were valid. However, it did provide encouragement to proceed with a validation.

Chapter 3 is a technical paper which investigated the difference in results produced by the original and newer version (v1.2.2) of STARFM, providing guidance for future research. As
previously noted, Chapter 2 identified some variation in the results between the two software versions. However, the improvements were not clear and since the paper in Chapter 2 was published a small bug has been fixed in the software. To allow future use of STARFM v1.2.2, Chapter 3 validated the new version and compared it to the original version, concluding that the new version should be adopted in future work.

Chapter 4 built on Chapters 2 and 3. By using the results of Chapter 3, v1.2.2 was run to produce a time series of fused images over three rural to urban gradients. Phenophases were then extracted from the time series and compared to the locations of in situ observations of the USA-NPN for the leaves phenophases of the red maple (*Acer rubrum*). The results showed that usually the two data sources were in agreement that spring is advanced in the urban areas and that some USA-NPN phenophases in some cities had a statistically significant relationship with the phenophases based on those fused images, but that most of the variance was left unexplained. In addition, the average difference in the urban effects on phenology from the fused data and actual red maple phenology varied from city to city.

Chapter 5 provided a brief synthesis of this entire document. It tied together the main findings of Chapters 2-4 and identified remaining gaps and limitations.
7. References:


30


Yuan, F., & Bauer, M. E. (2007). Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat


Chapter 2

Modeling the Effects of the Urban Built-Up Environment on Plant Phenology Using Fused Satellite Data

*The contents of this chapter have been previously published under a Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/) and were then reformatted for this dissertation. The citation to the published article:


Norman Gervais 1,*, Alexander Buyantuev 2 and Feng Gao 3

1Department of Information Science, University at Albany, State University of New York, New York, NY 12222, USA

2Department of Geography and Planning, University at Albany, State University of New York, New York, NY 12222, USA; abuyantuev@albany.edu

3Hydrology and Remote Sensing Laboratory, Agricultural Research Services, United States Department of Agriculture, Beltsville, MD 20705, USA; feng.gao@ars.usda.gov

*Correspondence: ngervais@albany.edu; Tel.: 518-956-8246

Academic Editors: Naser El-Sheimy, Zahra Lari, Adel Moussa, Roberto Colombo, Qihao Weng, Clement Atzberger and Prasad Thenkabail

Received: 1 September 2016; Accepted: 17 January 2017; Published: 23 January 2017
Abstract: Understanding the effects that the Urban Heat Island (UHI) has on plant phenology is important in predicting ecological impacts of expanding cities and the impacts of the projected global warming. However, the underlying methods to monitor phenological events often limit this understanding. Generally, one can either have a small sample of in situ measurements or use satellite data to observe large areas of land surface phenology (LSP). In the latter, a tradeoff exists among platforms with some allowing better temporal resolution to pick up discrete events and others possessing the spatial resolution appropriate for observing heterogeneous landscapes, such as urban areas. To overcome these limitations, we applied the Spatial and Temporal Adaptive Reflectance Model (STARFM) to fuse Landsat surface reflectance and MODIS nadir BRDF-adjusted reflectance (NBAR) data with three separate selection conditions for input data across two versions of the software. From the fused images, we derived a time-series of high temporal and high spatial resolution synthetic Normalized Difference Vegetation Index (NDVI) imagery to identify the dates of the start of the growing season (SOS), end of the season (EOS), and the length of the season (LOS). The results were compared between the urban and exurban developed areas within the vicinity of Ogden, UT and across all three data scenarios. The results generally show an earlier urban SOS, later urban EOS, and longer urban LOS, with variation across the results suggesting that phenological parameters are sensitive to input changes. Although there was strong evidence that STARFM has the potential to produce images capable of capturing the UHI effect on phenology, we recommend that future work refine the proposed methods and compare the results against ground events.

Keywords: urban heat island (UHI); phenology; STARFM; remote sensing; TIMESAT; growing season
1. Introduction

The world’s urban population has been growing quickly from only 30% of the global population residing in urban areas in the 1950s to as much as 54%, or 3.9 billion urban residents, in 2014. It is projected to increase toward 66% by 2050 (United Nations, 2014), so it is important to understand the effects of urban areas on the environment. One such effect is the elevated temperature in urban areas, which in turn affects the growth cycle of plants known as plant phenology.

Urban heat island (UHI) is the term used to describe higher temperatures in an urban environment compared to the surrounding rural areas (Oke, 1982). UHI is mainly a result of changing a natural landscape into a man-made urban texture (Luo & Asproudi, 2015) and arises due to the modified surface affecting the storage and transfer of heat, water, and airflow (Hartmann et al., 2013). The UHI has a measurable effect, as previous research has shown a small UHI effect of 0.17 °C for cities as small as 2–3 km² and a larger effect of 2.9 °C during the day and 2.3 °C during the night for larger cities of >100 km² (Tan & Li, 2015). Since the UHI is comparable to projected global temperature increases (Jochner & Menzel, 2015), urban environments have been used as a surrogate for future climate change (i.e., Ziska et al., 2003).

The relationship between urban intensity and the course of annual developmental events in plants, known as plant phenology, has been documented (Walker, de Beurs, & Henebry, 2015). From a biological perspective, phenological research predominantly addresses the timing of switches between recurrent phases of organisms (Badeck et al., 2004). In cool and temperate regions, to maximize growth, trees need to extend the growth period as long as possible while avoiding frost damage (Häkkinen, Linkosalo, & Hari, 1998). It has been confirmed that bud burst is under strong temperature control, with leaves appearing earlier and growing faster during a
warmer spring (Wesołowski & Rowiński, 2006). In the fall, growth cessation and dormancy are in response to the shortening of the photoperiod (Perry, 1971), however, low temperature also causes growth cessation (Heide, 1974). Plants, in general, will develop earlier in cities by a factor of days to weeks when compared to their rural settings. However, there is no clear association with autumn phenophases. Comparing the urban to rural trends of phenology allows for the evaluation of phenological trends in differing conditions and assessment of impacts of climate change (Jochner & Menzel, 2015).

Earlier greening of land cover and later autumnal leaf fall may alter seasonal climate through biogeochemical processes and physical properties (Peñuelas, Rutishauser, & Filella, 2009). Phenological changes will have varying impacts on albedo (Richardson et al., 2013) and surface temperature (Schwartz & Karl, 1990). In addition, the length of the growing season plays a key role in the amount of CO2 in the atmosphere (Peñuelas & Filella, 2001). Earlier spring onset leads to an increase in Gross Primary Production (GPP) and ecosystem respiration, with GPP having the larger increase resulting in a higher Net Ecosystem Productivity (NEP) (Richardson et al., 2009). From a biological perspective, due to species responding differently to warming temperatures, mismatches in the phenology across trophic levels may already be occurring due to non-uniform shifting of co-occurring organisms, changing the biological communities (Primack et al., 2009).

Satellite derived measurements of land surface reflectance have been used to study land surface phenology (LSP) over large areas (for example, de Beurs & Henebry, 2004). When applied to examine the length of growing season for urban and rural areas, it was found that the growth season was about 15 days longer in urban areas (Zhang, Friedl, Schaaf, Strahler, & Schneider, 2004). However, this type of analysis has historically been restricted due to clouds,
atmospheric conditions, and the infrequent satellite coverage (Holben, 1986). Removing these atmospheric effects has a relatively long history (S. Liang, Fang, & Chen, 2001). However, even after this correction, the level of detail that urban maps provide is dependent on the spatial resolution of the sensor (Van de Voorde, Jacquet, & Canters, 2011) and the error in phenology detection increases with a reduction in temporal resolution (Zhang, Friedl, & Schaaf, 2009). Therefore, both high temporal and high spatial satellite sensors each have complementary advantages when observing urban phenology.

With the goal to obtain more information than each sensor can independently provide (Pohl & Van Genderen, 1998), a number of blending algorithms have been developed that focus on spatial and temporal dynamics (Emelyanova, McVicar, Van Niel, Li, & van Dijk, 2013). One such method is the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM). It was designed to predict surface reflectance with the spatial resolution of Landsat and the temporal resolution of the Moderate Resolution Imaging Spectroradiometer (MODIS) (Gao, Masek, Schwaller, & Hall, 2006). STARFM is, as of 2013, widely used (Emelyanova et al., 2013). Originally shown to capture phenology changes for forested or cropland areas, as well as a mixture of cropland, evergreen, and deciduous forest (Gao et al., 2006), it has since been applied to observe, for example, community level phenology (L. Liang et al., 2014), crop growth stages (Gao et al., 2017), urban phenomena such as the land surface temperature (LST) of the urban heat island (Shen, Huang, Zhang, Wu, & Zeng, 2016), and the Normalized Difference Vegetation Index (NDVI) values of a semi-arid rangeland of northern Utah (Olexa & Lawrence, 2014).

This study aims to build on the existing research and use blended, also known as fused or synthetic, satellite imagery to examine how the built up environment affects LSP of developed
land in a metropolitan area of Northern Utah. We derived 12 years of fused Landsat and MODIS reflectance data using the STARFM method with three different selection criteria for input data. Then, from the resulting fused datasets, we derived three time-series of a standard vegetation index for the study area to estimate phenology metrics and make comparisons of the urban and exurban area. Our study addressed three important research questions:

1. In our study area, is there an observable difference in key phenological parameters (Start of Season (SOS), End of Season (EOS), and Length of Season (LOS)) between the urban and exurban areas when using fused imagery over several years and how sensitive those differences to base pair selection?

2. If there is an observable difference in SOS, EOS, or LOS between the urban and exurban areas when using fused imagery, how often do these differences appear?

3. Does the amount of impervious area at a location affect any observable differences in SOS, EOS, or LOS between the urban and exurban areas when using fused imagery?

2. Materials and Methods

2.1. Study Area

The study area falls toward the southern edge of Landsat Path/Row 38/31, starting at about 35 km north of Salt Lake City, Utah (Figure 2-1). The urban and exurban areas were classified as such based on the average percent developed imperviousness for each municipality. These values were derived from performing a mean Zonal Statistic on the NLCD 2011 Percent Imperviousness (2011 Edition, amended 2014) data (available at MRLC, n.d.) with the zones defined by the geospatial dataset “Municipal Boundaries” found on the UTAH Automated Geographic Reference Center (available at: 

44
https://gis.utah.gov/data/boundaries/citycountystate/). The results showed a natural break in percent impervious surface, with no municipalities having a mean percent impervious surface between 23% and 40%. This break was used to define the urban and exurban areas and is similar to the criteria used by Imhoff et al. (Imhoff, Zhang, Wolfe, & Bounoua, 2010). They saw that, for most of the largest biomes in the continental US except desert and xeric shrubland ecoregions, urban areas of 25% impervious surface area (ISA) or greater have a higher average LST than suburban and rural areas of less than 25% ISA. This classification resulted in the municipal boundaries of areas that had an average of 41%–51% imperviousness representing the urban area and included Clearfield, Ogden City, Roy City, South Ogden, Sunset, and Washington Terrace. The boundaries of areas that had an average of 5%–22% imperviousness represented the exurban areas and included West Point, West Haven, Marriott-Slaterville City, Plain City, and Farr West (Figure 2-1). We used the municipal boundaries dataset to calculate geometry for each municipality in ArcGIS and revealed a total of 129.3 km² for the urban areas and 110.9 km² for the exurban areas (Table 2-1).

Table 2-1. Area of each municipality.

<table>
<thead>
<tr>
<th>Urban Municipal Name</th>
<th>Area (km²)</th>
<th>Exurban Municipal Name</th>
<th>Area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunset</td>
<td>3.8</td>
<td>West Point</td>
<td>18.5</td>
</tr>
<tr>
<td>Clearfield</td>
<td>20.0</td>
<td>West Haven</td>
<td>26.8</td>
</tr>
<tr>
<td>Roy</td>
<td>20.3</td>
<td>Marriott-Slaterville</td>
<td>19.1</td>
</tr>
<tr>
<td>Ogden</td>
<td>70.4</td>
<td>Plain City</td>
<td>31.2</td>
</tr>
<tr>
<td>Washington Terrace</td>
<td>5.2</td>
<td>Farr West</td>
<td>15.3</td>
</tr>
<tr>
<td>South Ogden</td>
<td>9.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>129.3</strong></td>
<td><strong>110.9</strong></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2-1. Location map of the study area within the contiguous United States and mean impervious surface and urban/exurban classification by municipality.

Table 2-2 shows the calculated mean urban and exurban LSTs for several leaf-off and leaf-on dates. To perform the calculation, we first converted the at-sensor spectral radiance for the thermal band 6 of selected Landsat 5 images into at-sensor brightness temperature and then calculated final surface temperatures by correcting for spectral emissivities according to (Artis & Carnahan, 1982). Emissivities were estimated using the NDVI thresholding method of Sobrino et al. (Sobrino, Jiménez-Muñoz, & Paolini, 2004), using NDVI values that were calculated from reflectance images downloaded from the Climate Data Record (CDR) collection available on the USGS’s EarthExplorer website. Clear LST data were extracted using a mask produced by the CFMask QA data provided with the Landsat reflectance data. Last, zonal statistics in ArcGIS were used to calculate the mean temperatures for the urban and exurban areas by date. The
results show that for 75% of the dates calculated, average LST is higher in the urban areas by a range of 1.38–2.98 °C. The other 25% of the areas show a difference less than 0.1 °C.

Table 2-2. Mean urban and exurban temperatures as well as the difference between them.

<table>
<thead>
<tr>
<th>Day</th>
<th>Exurban Temp. (°C)</th>
<th>Urban Temp. (°C)</th>
<th>Degrees Celsius Warmer Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>21, Year 2003</td>
<td>5.5</td>
<td>36.8</td>
<td>31.3</td>
</tr>
<tr>
<td>181, Year 2003</td>
<td>28.9</td>
<td>11.0</td>
<td>17.9</td>
</tr>
<tr>
<td>245, Year 2005</td>
<td>11.0</td>
<td>22.2</td>
<td>11.2</td>
</tr>
<tr>
<td>58, Year 2005</td>
<td>22.2</td>
<td>8.4</td>
<td>13.8</td>
</tr>
<tr>
<td>298, Year 2007</td>
<td>8.4</td>
<td>28.1</td>
<td>20.7</td>
</tr>
<tr>
<td>48, Year 2007</td>
<td>28.1</td>
<td>30.2</td>
<td>2.1</td>
</tr>
<tr>
<td>128, Year 2007</td>
<td>12.5</td>
<td>29.9</td>
<td>17.4</td>
</tr>
<tr>
<td>208, Year 2007</td>
<td>29.9</td>
<td>33.2</td>
<td>3.3</td>
</tr>
</tbody>
</table>

The elevation for these municipalities ranges from approximately 1286 to 1424 meters (USGS, n.d.). According to the National Weather Service, the local forecast office for Ogden, UT is Salt Lake City, UT. We used this local forecast office as an estimate to describe our study area’s conditions. From 1981 to 2010, Salt Lake City had a normal winter average temperature of slightly below freezing at −0.39 °C and a normal summer average temperature of 23.94 °C. The monthly average minimum temperature for this period ranged from −5.78 °C in January to 18.17 °C in July and the monthly average maximum temperature ranged from 3 °C to 33.67 °C, respectively. For this same timeframe, the normal water year precipitation was 409 mm with normal seasonal snowfall of 1427 mm (NOAA, n.d.).
2.2. STARFM Model Description

The theoretical basis for STARFM, as outlined by Gao et al. (Gao et al., 2006), was based on the premise that MODIS and Landsat surface reflectances are comparable, with small biases. Neglecting any small bias, the surface reflectance of a heterogeneous coarse pixel at date $t (C_t)$, is equal to the sum of each of the finer resolution homogenous pixel ($F$) times the percent of the area which the fine resolution pixel covers ($A$) at each location $i$, written as:

$$C_t = \sum (F_t^i * A_t^i)$$ (1)

For a MODIS image that has been super sampled to the equivalent fine resolution and boundary of a Landsat image with the same coordinate system, the homogenous MODIS course pixel ($M$) from the same location ($x_i, y_i$) and time ($t_0$) of the Landsat fine pixel ($L$) is equal to the Landsat pixel plus the difference between the observed MODIS and observed Landsat values, written as (Gao et al., 2006):

$$L(x_i, y_j, t_0) = M(x_i, y_j, t_0) + \varepsilon_0$$ (2)

As explained by Gao et al. (Gao et al., 2006), in an ideal situation, ground coverage and system errors for any given pixel does not change from a base date ($t_0$), when both coarse and fine resolution values are known for a pixel as described above, to the prediction date ($t_k$), when course resolution values are known but fine resolution values are unknown and need to be predicted. In the situation where there is no change, which is often not satisfied, $\varepsilon_0 = \varepsilon_k$ and:

$$L(x_i, y_j, t_k) = M(x_i, y_j, t_k) + L(x_i, y_j, t_0) - M(x_i, y_j, t_0)$$ (3)

To incorporate additional information from neighboring pixels, a central fine resolution pixel can be computed with the weighting function (Gao et al., 2006):

$$L(x_{w/2}, y_{w/2}, t_k) = \sum_{i=1}^{w} \sum_{j=1}^{w} \sum_{k=1}^{n} W_{ijk} \times (M(x_i, y_j, t_k) + L(x_i, y_j, t_0) - M(x_i, y_j, t_0))$$ (4)
where $w$ is the search window size and $L(x_{w/2}, y_{w/2}, t_k)$ is the pixel that is in the center of this search window. $W_{ijk}$ is a weight that determines how much each neighboring pixel contributes to the predicted reflectance of the central pixel. This weight is higher for neighboring pixel locations where the difference between the fine and course pixel of the same time is smaller, the difference between the coarse pixels at the base and prediction date are smaller, and the neighboring pixel is closer in distance to the central pixel. Within the search window, the algorithm uses spectrally similar pixels that are cloud free (Gao et al., 2006).

### 2.3. Satellite Data Processing

In accordance with Olexa and Lawrence’s (Olexa & Lawrence, 2014) validation of STARFM for the same Landsat Path/Row, Landsat TM data were used as the high spatial resolution images. The choice of high temporal resolution was based on previous findings (Walker, de Beurs, Wynne, & Gao, 2012) that favored the MODIS Bidirectional Reflectance Distribution Function (BRDF) adjusted 16-day reflectance (MCD43A4) product over MODIS 500 m daily surface reflectance (MOD09GA) data. Therefore, we chose to use the nadir BRDF-adjusted reflectance (NBAR) data with an eight day overlapping temporal resolution from the MODIS collection 5 data products.

USGS’s EarthExplorer website provided the download for the Landsat TM reflectance images from the Climate Data Record (CDR) collection. Two different methods were used to select the Landsat images used in the base pairs. The first used an informal screening process, similar to Olexa and Lawrence (Olexa & Lawrence, 2014), where we visually inspected Landsat scenes to ensure that there was relatively minimal cloud, shadow, snow, or water in our study area and immediately adjacent to it. To do this, we downloaded 80 images for path 38/row 31
with less than 10% cloud cover spanning the period from the start of the MODIS data production until the end of Landsat 5 (Day 49, 2000–Day 361, 2011). The CDR data also included identification for clouds, cloud shadows, snow, and water (USGS, 2015).

Python scripts were written to automate the process of preparing Landsat data for STARFM by looping through the images at each step, relying on the ArcPy library (ArcGIS Desktop: Release 10, n.d.) for many steps. The first step was to extract the files and rename the Red, Near Infrared (NIR), and Quality Assurance (QA) data for sorting and pairing when defining the STARFM inputs. The data were then reprojected to the WGS84 UTM_Zone12N coordinate system and clipped to the same extent. Based on the CFmask QA band that comes as a part of the CDR data, the scripts masked out all areas that represent fill, water, shadow, snow, and cloud, removing them from the analysis. Last, the scripts converted the images to a generic 16-bit signed binary file. Figure 2-2 shows a schematic that details the flow of the Python algorithm. Processed images were visually inspected under the informal screening criteria to remove those with high levels of no data values in and immediately adjacent to the study area, which left 54 Landsat images available for analyses. Removing a processed Landsat scene does not reduce the number of images used in determining seasonality parameters, instead it adds in an additional prediction day by removing a base pair.
Figure 2-2. The steps contained in the python script to process the MODIS and Landsat data for input into STARFM.

The second Landsat selection criteria used a statistical approach similar to Walker et al. (Walker et al., 2012) where the selection was based on the percent of unclear pixels. To do this,
for all available Landsat surface reflectance images we calculated the proportion of clear pixels within a 1500 m buffer of our study area, which represents the entire area that STARFM used for the search window. To determine the total number of clear pixels, we reclassified the clear flag in the CFMask file from the Landsat Surface Reflectance product so that clear pixels are set to a value of 1 and snow, water, cloud, and cloud shadows are set to 0. We then preformed zonal statistics on each reclassified CFMask that summed up all pixel values within our buffer area, considering no data pixels as not clear. To determine the total number of possible pixels, we created a raster image with 30 m resolution that encompassed the study area plus the 1500 m buffer and obtained a count of the pixels (504,714 in total). We then divided the total number of clear pixels by the total count and multiplied by 100. We set two statistical thresholds of the amount of clear pixels needed to include the image, one at 90% and the other at 95%. The same Python script then prepared all images that met either of these criteria for STARFM.

Then the NASA Reverb ECHO website provided all 546 NBAR (MCD43A4) images corresponding to the start of the MODIS data production through the end of the Landsat 5 data production. The BRDF describes the scattering of light from one direction into another (Martonchik, Bruegge, & Strahler, 2000). BRDF effects are not a concern when developing a surface reflectance product for Landsat imagery since it is acquired within less than 7.5 degrees of nadir (Masek et al., 2006). Due to no data being collected by MODIS from Day 166 to Day 184 in 2001 (NASA, 2016), the MODIS scene that corresponds to this timeframe was omitted, leaving us with 545 MODIS scenes to use. The Landsat scene from Day 175 of 2001, which paired with this MODIS scene, was also removed from analysis leaving 53 Landsat images available for analyses based on the informal screening process, 81 images based on a 95% threshold, and 92 images based on a 90% threshold (Figure 2-3).
Figure 2-3. The temporal distribution of the Landsat scenes used in the base pairing.

MODIS data were also prepared for STARFM using Python scripts. The scripts first extracted the Red and NIR bands and reprojected the images to the same coordinate system as Landsat. To help with reducing the file size before further processing, these data were subset to include a buffer around the final extent. The buffer allowed us to resample MODIS data to the same 30 m resolution as Landsat in a shorter amount of processing time while still preserving the ability to apply a final mask of the same extent of Landsat. If the final extent was applied before resampling, the edge of the image would include additional areas of no data due to the clip line bisecting a larger cell. Then the scripts shifted the cells to align perfectly with Landsat data. Last, all images were converted to a 16-bit signed generic binary format (Figure 2-2). The step of shifting cells was simplified for the MODIS data used in the statistical selection methods by shifting the final clip file instead.

2.4. STARFM Parameters and Input Text Creation

STARFM can either use one or two base pairs. Based on Olexa and Lawrence’s (Olexa & Lawrence, 2014) satisfactory results, we opted for one base pair to be used. In addition, to
coincide with Olexa and Lawrence (Olexa & Lawrence, 2014), the surface reflectance uncertainties were set to 0.002 for the Red and 0.005 for the NIR bands, and 40 for the spectral similarity test. However, we doubled the search distance to 1500 m to help compensate for the heterogeneous urban environment. As recommended by the software README file, the spatial flag was set to on. STARFM requires an input text file that specifies all of these input parameters and the names of the base pair images and the predicted image. Two custom Python scripts automated this, one for creating synthetic images for days which do not coincide with a base pair, and one for creating synthetic images for days which do coincide with a base pair. Three separate sets of input text files were created for each base pair selection method.

The overall approach for these scripts first identified the base pairs for the Red bands, or the Landsat and MODIS scenes that were closest to each other in date. To do this, the scripts compared the middle date (8th day) of the 16 day MODIS acquisition time to the Landsat acquisition date. If the difference between the Landsat and MODIS dates was between $-4$ and $3$, this signified that the MODIS scene had the middle acquisition day closest to the Landsat acquisition date and the two images should be a base pair. If the difference was outside of this range, the MODIS scene became a coarse resolution image that had a high resolution image predicted for it. For each predicted scene identified, these scripts created a text file with a unique name, a unique name for the predicted high resolution image was specified, and the name of the MODIS scene for the predicted day was written to the text file. Next, the names of the base pairs that had the Landsat acquisition date closest to the middle acquisition day of the predicted MODIS scene were added to the text file as base pairs. Last, the file had all constant input parameters written to it. The scripts then copied the file, changed the names from representing Red to NIR images, and the process was repeated for the next MODIS image (see Figure 2-4).
A similar script created the input files for prediction days that coincide with base pair dates. The reason for this prediction is mainly to have a 30 m file with the same name as the rest of the predicted days, but also to account for and apply any underlying biases that may or may not have existed in the methodology in a uniform manner. The main differences are that during the initial pairing, the predicted day assumed the MODIS name as well as the base pair day.

STARFM software version 1.1.2 was then run by looping through the input texts for both the informal selection criteria and the 95% threshold criteria. The 90% threshold criteria was processed with STARFM version 1.2.1. Version 1.2.1 improves computational efficiency by allowing predictions for multiple dates with one run when the pair image is the same. It can also run in parallel computing mode (Gao et al., 2017). STARFM can fill small gaps or clouds in the prediction MODIS image using information from neighbor pixels if Landsat pixels are valid. These predictions are useful but may be less accurate. The version 1.2.1 constrains the predictions for the cloud pixels in the MODIS pair and prediction images. Predictions for clear pixels in Landsat and MODIS images are the same from both versions.
Figure 2-4. The steps contained in the python script to match MODIS scenes without a corresponding Landsat scene with the appropriate MODIS/Landsat base pair and create the STARFM input texts reflecting this.
2.5. NDVI Time Series Preparation and Seasonality Extraction

Vegetation indices (VI) are correlated to photosynthetically active radiation absorbed by plants in the visible spectrum (Tucker, 1979). Normalized Difference Vegetation Index (NDVI) is the most widely used VI (Myneni, Hall, Sellers, & Marshak, 1995). This NDVI signal allows us to estimate seasonal changes in light absorption of a surface (Badeck et al., 2004) and may summarize the annual and interannual variability of plant phenology (Marchetti, Minotti, Ramonell, Schivo, & Kandus, 2016). The fused images produced by STARFM were used to calculate NDVI values (for example, see Figure 2-5) at an eight-day time series interval for each of the three methods with the following formula (Myneni et al., 1995):

\[
NDVI = \frac{NIR - RED}{NIR + RED}
\]  

(5)

Figure 2-5. Example synthetic NDVI image.

This produced a range of values from $-1$ to $1$, where negative values were related to areas with open water and positive values were areas covered by green vegetation (Marchetti et al., 2016). All locations with no data in either of the synthetic images were set to be well outside of
this range. For the time series produced by the statistical methods, areas where the Landsat image used in the base pair had a value for fill, water, shadow, snow, or cloud in the CFMask were also set to be well outside of this range. Based on the three NDVI time series values, the SOS, EOS, and LOS were all extracted using TIMESAT 3.2 software.

TIMESAT is an accurate method for extracting seasonality parameters (Jönsson & Eklundh, 2002). The SOS and EOS were defined as points in time where the value has increased or decreased by a certain amount (Jönsson & Eklundh, 2004). We used the following settings for this study: no spike method, logistic function for data smoothing, one envelope iteration, an amplitude season start method, an adaptation strength of two, and the amplitude value of 0.2 (or 20%) for both SOS and EOS. For specific details on these parameters, please consult the TIMESAT manual (Eklundh & Jönsson, 2015). In addition to using the calculated NDVI values in TIMESAT, since according to the software manual TIMESAT requires the same number of images for each year, a dummy image with values well outside of the NDVI value range was created and used for the day when no MODIS data was collected and every eight days between Day 1 and Day 41 of 2000. In addition, to ensure TIMESAT extracted data for all of the years in our study, two dummy years were included in the TIMESAT input that were duplicates of the first and last years of data and placed at the beginning and end of the time series. During seasonality extraction, the date range of synthetic images that specified when each SOS was expected to occur included three images (time steps) before the start of the year through the end of the year (i.e., image range 44–92 was used for the year 2000 which spanned images 47–92), for EOS the date range included the start of the year to three images after the end of the year (i.e., image range 47–95 for the year 2000), and the date range for LOS included three images
before the start of the year through three images after the end of the year (i.e., image range 44–95 for the year 2000).

### 2.6. Dates of Phenological Events for Developed Areas

Start of Season (SOS), End of Season (EOS), and Length of Season (LOS) were estimated for developed areas in the study area based on a total of 177,480 pixels representing the Developed Land Cover Classes (21 (Developed, Open Space), 22 (Developed, Low Intensity), 23 (Developed, Medium Intensity), and 24 (Developed, High Intensity)) within the 2011 NLCD Land Cover (amended 2014) data (available at MRLC, n.d.). For both SOS and EOS, the day of the event for each year was calculated by subtracting the number of images that were from years prior and then multiplying the result by the time step of eight days (Eklundh & Jönsson, 2015). Since the LOS output provided the time of the length and not a date, multiplying the length time by eight days produced the length in days.

### 2.7. Statistical Analyses

The raw dates for the urban and exurban areas provided us with the data for basic descriptive statistics. Similar to Zipper et al. (Zipper et al., 2016), prior to any comparisons that aggregated the years together all data were normalized to account for different temperature-based seasons. We normalized for each phenological parameter by year according to the expected exurban value for each land class by subtracting the mean value of the exurban area from each data point for each year. The resulting normalized values represented the number of days between the expected timing of the event in the exurban area and the timing of the event at each given point. Therefore, the time averaged zero for the exurban area and the urban area’s average
represented the difference in days of the event between the urban and exurban area. After normalization, each land cover class was extracted according to if it is within the exurban or urban areas and then stacked using RStudio (RStudio Team, 2015).

To help determine the appropriate statistical tests to run, we tested the stacked datasets for normality by visually inspecting a histogram of the data. Then the data were tested for equal variances between each urban and exurban sample using an F-test. We then performed a t-test to test if the mean of the urban area is statistically different from the mean of the exurban area.

3. Results

3.1. Accuracy Assessment of STARFM

Although our study area falls into the area of a previous study (Olexa & Lawrence, 2014) which validated STARFM’s ability to fuse images in our general location, a brief accuracy assessment was performed to ensure the heterogeneous urban area of our study did not significantly alter the accuracy of synthetic Landsat images. To assess the accuracy, we compared reflectance values of several Landsat surface reflectance images to the synthetic Landsat scene that was closest in date. The comparison was made by reporting both the $R^2$ values and the mean absolute difference (Table 2-3). None of these Landsat images that were used in validation were used as a part of a base pair in STARFM. In addition, the scenes chosen represent both leaf-on conditions and leaf off conditions.
Table 2-3. Accuracy assessment results showing $R^2$ values and the mean absolute difference between synthetic NDVI values and Landsat surface reflectance NDVI values that were not used in data fusion.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Abs. Diff. All Areas</td>
<td>0.07</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
<td>0.07</td>
<td>0.12</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Mean Abs. Diff. Urban Locations</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
<td>0.06</td>
<td>0.06</td>
<td>0.11</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Mean Abs. Diff. Exurban Locations</td>
<td>0.09</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>$R^2$ All areas</td>
<td>0.59</td>
<td>0.85</td>
<td>0.87</td>
<td>0.72</td>
<td>0.72</td>
<td>0.36</td>
<td>0.81</td>
<td>0.89</td>
</tr>
<tr>
<td>$R^2$ Urban Locations</td>
<td>0.66</td>
<td>0.91</td>
<td>0.92</td>
<td>0.77</td>
<td>0.8</td>
<td>0.41</td>
<td>0.84</td>
<td>0.91</td>
</tr>
<tr>
<td>$R^2$ Exurban Locations</td>
<td>0.45</td>
<td>0.69</td>
<td>0.8</td>
<td>0.63</td>
<td>0.54</td>
<td>0.28</td>
<td>0.71</td>
<td>0.85</td>
</tr>
<tr>
<td>Mean Abs. Diff All Areas</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.11</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Abs. Diff Urban Locations</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.11</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Abs. Diff Exurban Locations</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.11</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ All areas</td>
<td>0.68</td>
<td>0.88</td>
<td>0.87</td>
<td>0.46</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ Urban Locations</td>
<td>0.72</td>
<td>0.9</td>
<td>0.92</td>
<td>0.54</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ Exurban Locations</td>
<td>0.57</td>
<td>0.82</td>
<td>0.76</td>
<td>0.35</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2. Descriptive Statistics

Since SOS dates for many observations occurred before 23 February 2000, when the MCD43A4 data first became available, we removed this year from further analyses of SOS and
LOS. In addition, since the descriptive statistics placed the EOS dates for many observations after the last year (2011) of data, further analysis for both EOS and LOS did not include 2011. For each year, we sampled 48,942 exurban and 128,539 urban points, although the number of samples used for any given year varied due to no data values. The maximum number of no data values for all three base pair selection methods for the SOS was between 30,489 and 33,005 for the entire exurban area in and 91,482 and 116,354 for the urban areas.

The normalized data of the informal selection show an earlier SOS for urban areas by a factor of 2.21 to 6.73 days on average, depending on the class. The more developed classes had the biggest difference in SOS and the least developed classes along with all of the classes together had the smallest difference. Although the developed areas as a whole and individual classes both show a later EOS for the urban areas, the number of days between the mean EOS for the urban and the exurban area had a lower range that was from 0.04 to 0.71 days. Lastly, the difference in the average length of season was longer in urban areas and ranges from being 2.84 days for all developed areas up to 6.8 days longer for the most developed areas, with the difference increasing for each class with percent imperviousness (Figure 2-6a). Standard deviations for the years were generally largest for the LOS parameter (Table 2-4).

The normalized data of the 95% statistical selection threshold show an earlier SOS for urban areas by a factor of 1.05 to 3.56 days on average. Although the developed areas as a whole and the two classes of developed open space and developed medium intensity show a later EOS for the urban areas up to 1.25 days, the other two classes show a later EOS for the exurban area of 0.35 and 0.79 days. Again, the difference in the average length of season was longer in urban areas and ranges from being 0.78 days up to 4.32 days longer. The difference for each class does not consistently increase with percent imperviousness (Figure 2-6b).
The normalized data of the 90% statistical selection threshold on STARFM version 1.2.1 also show an earlier SOS for all urban areas ranging from 0.28 to 3.55 days. All developed areas as a whole and all four classes show a later EOS for the urban areas ranging from 2.84 to 5.15 days. Again, the difference in the average length of season was longer in urban areas and ranges from being 3.22 and 7.02 days longer (Figure 2-6c).

**Figure 2-6.** Three bar charts showing the approximate average number of days the SOS is earlier, EOS is later, and LOS is longer in the urban area for (a) the informal selection criteria, (b) the 95% statistical criteria, and (c) the 90% statistical selection criteria. Error bars represent a 99.9% confidence interval.
Table 2-4. Standard deviations for the normalized data.

<table>
<thead>
<tr>
<th>Informal Base Pair Selection</th>
<th>All Developed</th>
<th>Open Space</th>
<th>Low Intensity</th>
<th>Medium Intensity</th>
<th>High Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOS Urban</td>
<td>18.4</td>
<td>16.6</td>
<td>16.3</td>
<td>17.7</td>
<td>21.7</td>
</tr>
<tr>
<td>SOS Exurban</td>
<td>17.7</td>
<td>16.0</td>
<td>16.4</td>
<td>17.6</td>
<td>22.9</td>
</tr>
<tr>
<td>EOS Urban</td>
<td>17.6</td>
<td>17.9</td>
<td>16.1</td>
<td>16.5</td>
<td>22.0</td>
</tr>
<tr>
<td>EOS Exurban</td>
<td>19.7</td>
<td>19.3</td>
<td>18.6</td>
<td>20.2</td>
<td>25.3</td>
</tr>
<tr>
<td>LOS Urban</td>
<td>24.5</td>
<td>24.9</td>
<td>22.3</td>
<td>23.3</td>
<td>29.9</td>
</tr>
<tr>
<td>LOS Exurban</td>
<td>26.2</td>
<td>25.8</td>
<td>24.5</td>
<td>25.9</td>
<td>35.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>95% Clear Base Pair Selection</th>
<th>All Developed</th>
<th>Open Space</th>
<th>Low Intensity</th>
<th>Medium Intensity</th>
<th>High Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOS Urban</td>
<td>19.0</td>
<td>16.4</td>
<td>16.8</td>
<td>18.6</td>
<td>26.0</td>
</tr>
<tr>
<td>SOS Exurban</td>
<td>19.2</td>
<td>17.5</td>
<td>18.4</td>
<td>19.8</td>
<td>22.3</td>
</tr>
<tr>
<td>EOS Urban</td>
<td>19.1</td>
<td>20.1</td>
<td>17.6</td>
<td>17.9</td>
<td>23.7</td>
</tr>
<tr>
<td>EOS Exurban</td>
<td>21.7</td>
<td>21.7</td>
<td>20.8</td>
<td>22.4</td>
<td>24.1</td>
</tr>
<tr>
<td>LOS Urban</td>
<td>26.1</td>
<td>25.7</td>
<td>23.7</td>
<td>25.2</td>
<td>34.0</td>
</tr>
<tr>
<td>LOS Exurban</td>
<td>29.4</td>
<td>29.5</td>
<td>28.5</td>
<td>29.9</td>
<td>31.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>90% Clear Base Pair Selection</th>
<th>All Developed</th>
<th>Open Space</th>
<th>Low Intensity</th>
<th>Medium Intensity</th>
<th>High Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOS Urban</td>
<td>28.3</td>
<td>22.7</td>
<td>27.3</td>
<td>28.8</td>
<td>35.5</td>
</tr>
<tr>
<td>SOS Exurban</td>
<td>24.6</td>
<td>22.0</td>
<td>24.3</td>
<td>26.1</td>
<td>29.4</td>
</tr>
<tr>
<td>EOS Urban</td>
<td>24.4</td>
<td>26.3</td>
<td>23.8</td>
<td>22.9</td>
<td>30.4</td>
</tr>
<tr>
<td>EOS Exurban</td>
<td>27.3</td>
<td>27.3</td>
<td>26.3</td>
<td>28.4</td>
<td>31.9</td>
</tr>
<tr>
<td>LOS Urban</td>
<td>36.8</td>
<td>36.6</td>
<td>35.9</td>
<td>36.5</td>
<td>47.3</td>
</tr>
<tr>
<td>LOS Exurban</td>
<td>36.9</td>
<td>36.5</td>
<td>35.8</td>
<td>38.6</td>
<td>42.8</td>
</tr>
</tbody>
</table>

To see how consistent these overall findings were from year to year between classes, one can also look at the 11 years separately. This showed that the average urban SOS date was earlier for up to all of the 11 years for developed high intensity urban areas when using informal base pair selection criteria, but also earlier in only six years when using 90% selection criteria for developed low intensity areas. The average EOS date was later for the urban areas from as few as three out of the 11 years to as many as nine, with all three selection methods showing the least amount of years in the developed high intensity class. The average LOS was longer in the urban areas for 10 out of 10 years for developed open space with an informal selection and for only six years for various classes with both the 90% and 95% selection criteria (Table 2-5).
Table 2-5. Frequencies of years where mean SOS was earlier, EOS was later, and LOS was longer in the urban areas according to base pair selection.

<table>
<thead>
<tr>
<th>Years with Earlier Urban SOS</th>
<th>Informal Selection</th>
<th>95% Selection</th>
<th>90% Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Developed Areas</td>
<td>9</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Developed Open Space</td>
<td>9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Developed Low Intensity</td>
<td>8</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Developed Medium Intensity</td>
<td>10</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Developed High Intensity</td>
<td>11</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Years with Later Urban EOS</th>
<th>Informal Selection</th>
<th>95% Selection</th>
<th>90% Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Developed Areas</td>
<td>7</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Developed Open Space</td>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Developed Low Intensity</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Developed Medium Intensity</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Developed High Intensity</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Years with longer Urban LOS</th>
<th>Informal Selection</th>
<th>95% Selection</th>
<th>90% Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Developed Areas</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Developed Open Space</td>
<td>10</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Developed Low Intensity</td>
<td>9</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Developed Medium Intensity</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Developed High Intensity</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

3.3. Tests for Variance Equality and Normality

Based on the results of the F-tests, almost all land cover classes and the developed areas as a whole had statistically unequal variances between the urban and exurban areas. The exceptions to this were the informal selection developed medium intensity land cover class for the SOS and the 90% selection criteria’s LOS for all developed areas, developed open space and developed low intensity, where all \( p \)-value were above 0.1. Visual inspection of all histograms showed an approximate normal distribution (see Supplementary Materials for the histograms). This method was chosen over a statistical test due to small deviations from normality being marked as significant even though they should not affect parametric test results (Öztuna, Elhan, & Ersöz Tüccar, 2006).
3.4. Tests for Inequality of Phenological Parameters

RStudio performed a two-tailed and two-sample t-test for each parameter and each developed area as a whole as well as each land cover class separately to see if the mean of the urban area was statistically different from the mean of the exurban area. The t-test assumed unequal variances in all cases except when the F-test could not show an unequal variance. The tests showed that for the SOS, both the informal and 95% selection criteria show a statistical difference between the urban and exurban areas at the 99.9% level for all areas together and all classes separately, with all observed \( p \)-values of \(<2.2 \times 10^{-16}\). However, after applying a Bonferroni correction, the 90% selection criteria only shows a 99.9% significant difference for all developed areas, developed medium intensity, and developed high intensity areas, with \( p \)-values of \(<2.2 \times 10^{-16}\) for all developed areas and medium developed areas and \(4.91 \times 10^{-15}\) for developed high intensity areas. It also shows an insignificant difference in the SOS for both developed open space and developed low intensity with \( p \)-values of 0.0045 and 0.013, respectively.

The EOS results were all statistically significant at the 99.9% level, with all observed \( p \)-values of \(<2.2 \times 10^{-16}\), for the 90% selection criteria. However, they were mixed for both the informal and 95% selection criteria. With the informal selection criteria, both developed areas as a whole and developed open space had an observed \( p \)-value of \(<2.2 \times 10^{-16}\). After applying a Bonferroni correction for the remaining classes, the results show no statistical difference for EOS in the developed low intensity, developed medium intensity, and developed high intensity classes with observed \( p \)-values of 0.448, 0.011, and 0.107, respectively. With the 95% selection criteria, both developed low intensity and medium intensity were statistically significant with \( p \)-values of \(6.5 \times 10^{-10}\) and \(<2.2 \times 10^{-16}\), respectively, and after the Bonferroni correction developed high
intensity was significant at the 99% level with an observed $p$-value of 0.000012. There was no significance for the remaining classes after a Bonferroni correction with all developed areas having a $p$-value of 0.03 and developed open space having a $p$-value of 0.75.

The length of season showed the most consistent results with most classes for all methods having an observed $p$-value of $<2.2 \times 10^{-16}$ making them significant at the 99.9% level. The only exceptions to this were the developed high intensity results for both the 95% and 90% selection methods. The 90% selection criteria results were still significant at the 99.9% level but the observed $p$-value was $1.5 \times 10^{-14}$. The 95% selection method still did show significance at the 99% level after a Bonferroni correction with an observed $p$-value of 0.000049.

4. Discussion

4.1. Start of Season

Our study analyzed the effects of UHI on phenology in urban areas, although we did not attempt to measure temperature for each land cover per se. Instead, we used percent imperviousness to understand the spatial variability of temperature in cities. Since UHI is the direct product of the transformation of natural landscapes into an impervious man-made urban texture (Luo & Asproudi, 2015) and a strong linear relationship between the percent impervious surface area and land surface temperature has been shown (Yuan & Bauer, 2007), it is reasonable to assume that the areas we defined as urban are warmer than the areas we defined as exurban. Based on previous findings suggesting that leaves appear earlier during a warmer spring (Wesołowski & Rowiński, 2006) and that leaves develop earlier in cities (Jochner & Menzel, 2015), we reasonably expected our urbanized areas to exhibit an earlier SOS, which was supported by the results of this study, although not statistically when using STARFM v1.2.1 and
a 90% selection criteria for all classes. Therefore, base pair selection and software version may make a difference when observing SOS of the urban environment. In addition, it has been shown that temporal gradients in leaf onset varies (Fisher, Mustard, & Vadeboncoeur, 2006), which coincides with our findings of a large standard deviation for SOS (Table 2-4).

Interestingly, our results do not show a similar magnitude of differences in the SOS between classes throughout the urban and exurban area. This coincides with the case study by Melaas et al. (Melaas, Wang, Miller, & Friedl, 2016), which showed that the amount of impervious surface area in surrounding vegetation patches influences the timing of SOS. Although we do not show a clear trend with the difference in SOS growing as mean percent imperviousness does, for all three selection methods the difference in mean days becomes larger between urban and exurban areas for the two most developed classes when compared to the two least developed classes. In addition, out of all 11 years where SOS was calculated, all methods showed an earlier urban SOS for all classes more times than a later urban SOS. However, there is a range to the number of years where urban SOS is earlier. Based on this, there is not a uniform effect on the SOS throughout the entire urban area. In addition, it is reasonable to assume that we may lose some information if we generalize entire municipalities as having an earlier SOS. For this reason, having a spatial resolution finer than that of MODIS, such as that produced by fused data, would allow for a better understanding of the complex relationships between heterogeneous urban and exurban environments.

4.2. End of Season

The EOS results were not as conclusive as the SOS results, which is not surprising since previous research also found an unclear association between the urban-exurban differences and
autumn phenophases (Jochner & Menzel, 2015) and high variability in the difference (White, Nemani, Thornton, & Running, 2002). Even though the mean EOS was later for urban areas in 13 out of 15 of the stacked datasets, the statistical significance across all classes was present only when the 90% criteria was used. This also indicates that base pair and software selection may lead to different results than when looking at features or specific locations within the municipality. This supports previous findings that EOS in urban vegetation patches is influenced by the percent impervious surface of surrounding patches (Melaas et al., 2016).

There are two potential reasons for such findings of EOS inconsistency between base pair selection and software. First, there could be a consistent and large difference in the EOS, as the 90% selection criteria with STARFM v 1.2.1 showed, but our methods were too simplistic or introduced excessive errors so the difference was improperly observed with using an informal and 95% selection criteria and STARFM v1.1.2. Second, there are potentially other environmental factors that are similar between the urban and exurban area, which are the main cues for trees to start growth cessation. For example, Perry (Perry, 1971) points to the shortening of the photoperiod. If this were true, we would not expect the UHI to change EOS. In addition, this would mean that the difference in EOS shown when using the 90% criteria is a result of observing a sample of the EOS under control of another environmental factor such as, for example, photoperiod.

4.3. Length of Season

Although LOS is the only phenological parameter that is consistently statistically significant, we are cautious of these results. Both the SOS and EOS contribute to the LOS, and under an informal and 95% selection criteria the most reliable contributor to the statistically
longer growing season in the urban area is the earlier start in the spring. As shown in Figure 2-6a,b, the difference in the SOS for the urban and exurban area is proportionally larger than the EOS, thus contributing more to the difference in the LOS. In addition, even without statistically different EOS dates for all of the classes with these two methods, the difference in the LOS is still statistically significant for all classes. However, when we look at LOS derived from the 90% criteria, the opposite is true that without all statistically different SOS dates, the difference in the LOS is still statistically significant for all classes. Therefore, since there is uncertainty with why LOS is longer in the urban areas, it is difficult to support that it is true.

Moreover, when looking at the mean LOS for each class, a similar trend appears as we see with the SOS, which is that there is a difference in the LOS across classes. Once again, this observation suggests to look at intra-city features and not just the urban area as a whole to be able to describe better the effects of the urban environment on phenology.

4.4. Study Assumptions and Potential Limitations

The method we used for normalization ensures that all data obtained are used. It also corrects for interannual shifts in phenological parameters by placing the event at each location in relation to the expected date of the same exurban event for the same year. This allows for the comparison of the difference in time between an event at each location and the timing of the corresponding expected exurban area’s event and not a comparison of the actual dates, which may not be comparable from one year to the next due to changing weather patterns. Thus, after normalization, one can compare the urban and exurban data to show the differences within each separate year and also among all years together. Importantly, the approach accounts for the uneven and inconsistent sample proportions between the urban and exurban environments from
year to year. Simply normalizing according to the mean value of all urban and exurban pixels would pull the normalized value in different amounts, based on the proportional size, toward the sample with the larger size and not centering the normalized values at the same value from year to year. For example, if one year had one exurban point for every two urban points and the next year this ratio decreased to 1:4, the normalized values of the second year would be pulled closer to zero for the urban areas and further from zero for the exurban areas, making them incomparable.

However, an underlying assumption of this normalization method is that if a shift in dates does occur from one year to the next, the amount of the shift is the same for both the urban and exurban environment. If this assumption is violated, any disproportional interannual shifts would skew the results to either be over or under represented depending on if the urban sample size was large or small for that year.

A second underlying assumption of this study is that the expected vegetation type and management practices were similar between like land cover classes for both the urban and rural environments. For example, it was assumed that any influences a homeowner may have on phenology due to landscaping practices (i.e., watering the lawn in the spring) and vegetation transplanting (i.e., planting grass for a lawn) were just as likely to occur regardless if the individual lives in a rural and urban settings. This is an important assumption, as it has been observed that phenological traits can differ between vegetation types (Zipper et al., 2016). However, making comparisons between the urban and rural environments of developed areas with the same land cover classes controlled for this, as opposed to comparing dissimilar land cover classes that may have dissimilar landscape management techniques.
The third assumption is that changes in NVDI values accurately represent changes in phenological parameters. The current study used changes in NDVI as a proxy for growing season parameters. However, it is not well understood how well these values, derived from fused imagery, correspond to the actual ground events of SOS, EOS, and LOS. The Collection 5 MODIS NBAR data used in this study were the 8-day overlapping product that is collected over 16 days. Depending on the data availability during the 16-day period especially around the SOS and EOS dates, we could miss or smooth the small changes from the multi-date data products.

The fourth limitation is that we did not attempt to resolve any potential errors with the NLCD classifications. Although overall accuracies were 79% and 78% for the previous Level II NLCD versions from 2001 and 2006, respectively, they did have difficulty distinguishing the context of grass. In relation to the land cover classes we looked at, the 2006 NLCD had a producer’s accuracy of 42%, 70%, 80%, and 26% for classes 21, 22, 23, and 24, respectively. The user’s accuracies for these classes were 52%, 59%, 69%, and 83%, respectively. The 2001 NLCD accuracy was similar (within 5%) to the 2006 accuracies except for the producer’s accuracy for class 24, which was 81% (Wickham et al., 2013).

The last assumption is that our sample was representative of the conditions of the entire urban and exurban areas. We had a very large sample, so this is reasonable to assume. However, the areas of no data predictions could be biased due to some uniform trait and therefore underrepresented. For example, the proportion of the sample remaining toward the center of the urban and exurban areas was similar for the informal and 95% selection criteria. However, for the 90% selection criteria the central areas were disproportionally underrepresented (Table 2-6). Previous research (Zhang et al., 2004) has shown a relationship between the UHI and distance to urban core area and should be considered.
Table 2-6. Total original sample size for all developed areas based on base pair selection and the percent of the original sample remaining 500 m and 1500 m inward away from the urban and exurban edges.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Total Sample Size</th>
<th>Percent Remaining at 500 m Inward</th>
<th>Percent Remaining at 1500 m Inward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal Selection</td>
<td>SOS 1,241,726</td>
<td>56</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>EOS 1,681,156</td>
<td>56</td>
<td>13</td>
</tr>
<tr>
<td>&gt;95% Selection</td>
<td>SOS 1,263,499</td>
<td>56</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>EOS 1,632,057</td>
<td>57</td>
<td>13</td>
</tr>
<tr>
<td>&gt;90% Selection</td>
<td>SOS 511,615</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>EOS 634,573</td>
<td>44</td>
<td>5</td>
</tr>
</tbody>
</table>

In addition, selecting Landsat images with less than 10% cloud cover as an initial screening process could potentially omit additional dates of clear Landsat images for the study area, thus reducing the number of base pairs used. Based on previous research (Fu et al., 2015), the risk of omitting these scenes potentially leads to a less than optimal quality of fused image since there could have been a potential base date nearer to some of our predicted dates than what we used for prediction. Adding to this, using a visual inspection to remove Landsat images could introduce human error.

4.5. Future Work

Future research should attempt to identify the source of variation that caused inconsistent results, being cautious of how input data would affect their sample distribution and validity of predictions. Evaluating base pair selection should include not only ensuring a valid fused image, but also taking into account how the distribution will affect the location of no data and how this in turn affects TIMESAT. Since the difference in base pair selection method does not produce
large differences in the validity of our fused images, how the data and parameters can affect the
ability of TIMESAT to make predictions is important to consider.

The MODIS collection 6 now provides daily NBAR product at 500 m resolution which
could be utilized in data fusion. The BRDF corrected daily surface reflectance is another option
to use in the future. The daily MODIS products may be able to reduce uncertainties in extracting
phenological parameters. This study used Landsat 5 TM data for data consistency. Landsat 7
ETM+ data even with the Scan-Line-Corrector errors can still be used to increase the frequency
of high resolution data. Other recent high resolution data such as Landsat 8 and Sentinel-2 can be
included in the future. The Landsat and MODIS pair images may be gap-filled first before data
fusion to increase valid prediction. Different weights from original and the fused data sources
may be considered similar to the strategy used in mapping crop phenology (Gao et al., 2017).

The results of future work should also validate the predicted phenological dates derived
from STARFM synthetic imagery against in situ observations of phenological parameters for the
urban and exurban environments, preferably for a variety of land cover classes. In addition, our
study was confined to a very specific environment and may not be generalizable to a variety of
areas. Expanding the prediction to a variety of climates and cities would help to overcome this.
Last, the value of using a fusion model over one source of satellite data to predict the differences
in the urban and exurban growing seasons should be explored. Our next steps will be to attempt
to model LSP parameters with high resolution/low frequency and low resolution/high frequency
data alone. This will allow making conclusions about the amount of information gained from
fusing the data in regards to observing the effects of the UHI on phenology.

Beyond lacking in comparisons to other data sources and locations, the current study did
not aim to explore in depth how different land cover classes affect phenological events. Although
this paper described the difference in mean urban and exurban values among land cover classes, it was limited to a simple descriptive analysis. In addition, it may be interesting to see if there is a shift in calendar dates of phenological events for both urban and exurban environments, rather than just look at the shift in the number of days of the normalized difference between the two as we did.

5. Conclusions

This study has uncertainties, but it offers encouragement that STARFM can be sensitive enough to detect the change in phenological parameters of vegetation that we would expect to see because of the UHI effect in urbanized landscapes. Although base pair selection for image fusion did not strongly affect the validity of fused images, it did have a noticeable effect on the phenology results. We suspect that this is in part due to the changing distribution of where no data are available for TIMESAT to fit. As the number of base pairs increases, so will the potential locations of no data from Landsat that is masked out from a prediction. As we saw with disproportional distribution of no data in one instance, these no data points might not be randomly distributed and therefore skew any results. We believe that this could have skewed the results for the 90% base pair selection because we not only added in the potential for an additional 5% unclear pixel coverage, but also due to the software change. STARFM v1.2.1 added in a strict validation that may potentially not allow a prediction to be made in areas where it did before, thus compromising a lesser sample size for an increase in overall prediction quality and not covering up any bias distribution in no data.

Our statistical results are mixed, but generally, they support the notion that the UHI leads to an early start of growing season (SOS) and an overall increase in the length of the growing
season (LOS). Conversely, the results show that depending on the methods used to select base pairs and the software version used to make the predictions, the effects urban areas have on the timing of the end of the growing season (EOS) can fluctuate. By using the fusion method, we were able to work toward understanding some of the advantages and challenges of modeling the heterogeneous urban landscape while preserving temporal accuracy. Overall, it allowed us to not only see a difference in phenological parameters between urban and exurban areas, but also to see that regardless of fusion method, there is variability among phenological parameters according to the land cover class at 30 m resolution. Depending on the base pair selection and version of STARFM, the differences in mean earlier urban SOS between land cover classes within each selection criteria ranged from as many as 4.5 to as few as 2.5 days, the differences in mean EOS urban delay between land cover classes ranged from 0.67 to 2.3 days, and the differences in mean longer urban LOS ranged from 3.6 to 3.9 days. Without having the temporal and spatial resolution of a fused image, these intercity variations in phenological parameters might not be observable.

However, how well the changes detected correlate to actual in situ conditions and the actual timing of phenological events is unknown. Lastly, it has yet to be determined how much information that describes phenology is gained by fusing the data and to what extent errors are introduced into these estimations.

6. Supplementary Materials

The following are available online at www.mdpi.com/2072-4292/9/1/99/s1, Figure S1: Histograms of stacked data for (a1–c10) informal selection of base pairs; (d1–f10) 95% clear threshold for selection of base pairs; and (g1–i10) 90% clear threshold for selection of base pairs.
7. Acknowledgments

The work was partially supported by the NASA Science of Terra and Aqua program (NNH13ZDA001N-TERAQ). A. Buyantuev’s research was partially supported by the NSF-MRI award #1531511. We would like to thank Sun Liang, of the USDA Agricultural Research Services, for helping confirm that our data processing procedures were correct. The USDA is an equal opportunity provider and employer. We would also like to thank John Pipkin, SUNY Albany, for his discussions on statistical methods; Shiguo Jiang, SUNY Albany, for helping with converting file formats; and Jim Mower, SUNY Albany, for his guidance. Last, we would like to thank the editors and the anonymous reviewers for their suggestions.

8. Author Contributions

N.G. and A.B. conceived and designed the experiments; N.G., A.B. and F.G. contributed analysis tools; N.G. performed the data processing and analysis; A.B. calculated the LST; and N.G. wrote the first draft of the paper, with significant revisions provided by A.B. and F.G.

9. Conflicts of Interest

The authors declare no conflict of interest.
10. References


https://doi.org/10.1088/1748-9326/11/5/054023

https://doi.org/10.1067/mai.2003.53

© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).
Chapter 3

Examining the upgraded STARFM: Quality and Quantity of Predictions

Abstract: STARFM is a commonly used fusion method for combining different sources of satellite data, one with high temporal resolution and the other complementary source with high spatial resolution. Resulting fused images retain superior traits from both the input images. Since the original release, strict validation and the ability to perform parallel processing were introduced in the most recent STARFM v1.2.x. However, at the time of this release it was not clear how much of an improvement it offers and at what expense it is achieved in relation to the reduction of the number of locations predicted. This technical paper quantified these issues by comparing the results of version 1.1.2 and version 1.2.2. Results show quality and quantity of predictions are not compromised for the decrease in processing time.

1. Introduction

Remote sensing has a variety of applications (United States Geological Survey, n.d.-a), including the study of land surface phenology (LSP), which requires a time series of image data (i.e. de Beurs & Henebry, 2004). However, different satellite systems have different issues and challenges with images they provide (for example, see Kennedy et al., 2009). When planning a remotely sensed investigation, one has to consider various data with different spatial, spectral, temporal, and radiometric characteristics (Woodcock & Strahler, 1987). For example, the finest spatial resolution that MODIS products can have is 250m, but they are produced daily (United States Geological Survey, 2014), whereas Landsat 8 has a spatial resolution of 15-100m, depending on the band (United States Geological Survey, n.d.-b), but a temporal resolution of 16
days (United States Geological Survey, 2017a). Spectral characteristics of key bands of Landsat and MODIS systems are designed to be comparable, although Landsat has higher spatial and MODIS has higher temporal resolution.

A number of blending (data fusion) algorithms have been developed to combine complementary data sources, one with high spatial resolution and the other with high temporal resolution, and produce one image with both high temporal and high spatial resolutions that better captures spatio-temporal dynamics. One widely used blending algorithm is the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Emelyanova, McVicar, Van Niel, Li, & van Dijk, 2013). It was originally developed to predict daily surface reflectance at the Landsat resolution (30m), but the approach can also be used with other similar instruments (Gao, Masek, Schwaller, & Hall, 2006). The first released version was v1.1.2, with improvements since being developed by Dr. Feng Gao of the USDA-ARS Hydrology and Remote Sensing Laboratory (HRSL) (USDA-ARS, n.d.). STARFM v1.2.1 was released with improvements in computational efficiency (Gao et al., 2015). These improvements have been shown to reduce the time to make 30 predictions with a single input pair from 540 minutes with v1.1 to just 17 minutes with v1.2 (Gao et al., 2017). In addition, v1.2.1 added in a strict validation (Gervais, Buyantuev, & Gao, 2017) and the current version, v1.2.2 fixed a small bug in fill values (USDA-ARS, n.d.).

Although the difference in computational efficiency was described in Gao et al. (2017), it was not clear if there are any effects on the quality of fused images produced as a result of the newer version’s strict validation. Gervais et al., (2017) used v1.1.2 and v1.2.1 to predict and validate Normalized Difference Vegetation Index (NDVI) values for two of the same dates, at least one of which used the same base pairs, and found that generally v1.1.2 produced better results. However, it was not clear if the small bug fixed in the v1.2.2 release would produce more
accurate predictions than the original release. In addition, although no direct comparison can be made about the number of pixels predicted by each version from the results of Gervais et al.’s (2017) previous study because different base pair selection criteria were used as software versions were changed, the study revealed a reduced number of predictions with the new version, which warrants a closer look. Therefore, this paper looks at these two topics and answers the following research questions:

1. Compared to STARFM v1.1.2, does v1.2.2 produce higher quality NDVI images?
2. By how much does STARFM v1.2.2 reduce the number of pixels predicted, if at all, in the fused images compared to the original version?

2. Methods

2.1. Image Fusion

STARFM was developed to combine the high spatial resolution of Landsat with the high temporal frequency of MODIS and produce blended images with high temporal and high spatial resolutions. Although originally it was used to produce daily reflectance values from the 500m resolution MODIS daily surface reflectance product (MOD09GHK) and the 30m resolution Landsat 7 Enhanced Thematic Mapper (ETM+) data, the method is applicable to the similar Landsat 8 Operational Land Imager (OLI) data (Gao et al., 2006). It has also been shown to produce better results with the high temporal frequency input data of the 16-day composite MODIS Nadir Bidirectional Reflectance Distribution Function (BRDF) Adjusted Reflectance (MCD43A4) dataset (Walker, de Beurs, Wynne, & Gao, 2012). In light of this, the scan line corrector (SLC) failure of Landsat 7, and the launch of Landsat 8 since the development of the
software, this study used the Landsat 8 OLI surface reflectance and MODIS MCD43A4 V006 products to produce fused images for several validation dates.

For this study, base pairs (Landsat and MODIS images from the same time) were selected and data were preprocessed to be the same temporal resolution, extent, and format in a way similar to the statistical selection criteria documented in Gervais et al. (2017) with three minor modifications. First, reprojecting MODIS was done after it was resampled to 30m. Second, when preprocessing MODIS all ‘no data’ values were reclassified to an extremely low value prior to applying the size clip. This was needed to remedy a problem with further analyses if a MODIS image was completely filled with no data. The third modification adjusted the scripts for assessing pixel quality. Instead of the previously used CF mask band, the values of 322 and 386 from the new quality assurance band (pixel_QA), part of the standard Landsat 8 surface reflectance product, were used (U.S. Geological Survey, 2017). All Landsat images used were Tier 1 scenes, which are the Landsat scenes that have the highest data quality (United States Geological Survey, 2017b). Since V006 of MCD43A4 data is produced daily, the MODIS image used in the base pair had the date of its central day the same as the date of the Landsat image.

After preprocessing, STARFM v1.1.2 and v1.2.2 were run on the Ubuntu operating system (https://www.ubuntu.com/) with STARFM parameters set the same as in Gervais et al. (2017).

2.2. Validation

To test the improvements, clear images were fused and validated for the same study area as Gervais et al. (2017). This location included 11 municipalities focused around and including Ogden, UT (See Figure 1 of Gervais et al. (2017), boundaries available here:}
In the period 2014-2016 in this location, 2015 has most scenes that were >95% clear (Table 3-1). From this year, the images from DOY 86 and 266 were used as base pairs. Both of these images were approximately 99% clear and were the first and last images in this year to be this clear.

To obtain a sample, cells of a 30m raster that encompassed the study area were converted to points and then all points within the 11 municipalities were selected. Raster cell values for fused and actual Landsat data were then extracted to the selected points and attribute data for those points exported to a table and processed in RStudio (RStudio Team, 2015). These data were cleaned by removing all locations with the lack of information for prediction or validation. This involved removing all points for each predicted day where values outside of the expected data range were found for either of the base pair images, the MODIS prediction day image, the Landsat validation day image, or the predicted image. In addition, any unclear locations as defined by pixel_QA for either the prediction or validation days were masked out.

To assess image quality, pixel based comparisons of reflectance values for the red band, near infrared (NIR) band, and calculated NDVI values were performed between the fused and actual Landsat images of the same day. All Landsat images used in the comparison were not used in base pairs during prediction and spanned a variety of dates in the calendar year (Table 3-1). Both the mean absolute difference (MAD) and $R^2$ values were calculated, similar to other studies (for example, see Gervais et al., 2017 and Walker et al., 2012). In addition, the number of pixels with valid values in the synthetic images was counted for each of the dates to assess the number of predicted locations.
Table 3-1. Validation dates (asterisks denotes base pair)

<table>
<thead>
<tr>
<th>Percent Clear</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>22-Jan</td>
</tr>
<tr>
<td>97</td>
<td>23-Feb</td>
</tr>
<tr>
<td>99</td>
<td>27-Mar*</td>
</tr>
<tr>
<td>99</td>
<td>12-Apr</td>
</tr>
<tr>
<td>99</td>
<td>28-Apr</td>
</tr>
<tr>
<td>95</td>
<td>17-Jul</td>
</tr>
<tr>
<td>99</td>
<td>18-Aug</td>
</tr>
<tr>
<td>99</td>
<td>19-Sep*</td>
</tr>
<tr>
<td>96</td>
<td>22-Nov</td>
</tr>
</tbody>
</table>

3. Results

The mean difference in $R^2$ values between the two software versions across all of the predicted dates is near zero for both base pair dates. The maximum $R^2$ difference for all of the bands and calculated NDVI ranges from $1.1 \times 10^{-5}$ up to $6.0 \times 10^{-5}$ (Table 3-2). The difference in the MAD across all dates between the software versions is also near zero, with a maximum difference of less than 0.1 for the scaled reflectances of the red and NIR bands and less than 0.0001 for the calculated NDVI (Table 3-3). For the red, NIR, and NDVI both software versions produced the same number of pixels for every date for both base pairs.
Table 3-2. Reported $R^2$ results pertaining to quality.

<table>
<thead>
<tr>
<th></th>
<th>R$^2$ Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DOY  22  54  102  118  198  230  262  326</td>
</tr>
<tr>
<td></td>
<td>STARFM v1.2.2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STARFM v1.1.2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase R$^2$ for new STARFM</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STARFM v1.2.2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STARFM v1.1.2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase R$^2$ for new STARFM</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3-3. Reported MAD results pertaining to quality. Red and NIR values have a scale factor of 0.0001.

<table>
<thead>
<tr>
<th></th>
<th>MAD Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DOY 22 54 102 118 198 230 262 326</td>
</tr>
<tr>
<td>STARFM v1.2.2</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>182.14 139.68 101.59 156.27 283.78 225.35 208.31 172.58</td>
</tr>
<tr>
<td>NIR</td>
<td>270.63 198.42 157.19 229.78 405.18 346.69 320.77 297.40</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.07740 0.05485 0.03944 0.06334 0.10439 0.10115 0.10358 0.07582</td>
</tr>
<tr>
<td>STARFM v1.1.2</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>182.13 139.67 101.59 156.28 283.78 225.35 208.32 172.56</td>
</tr>
<tr>
<td>NIR</td>
<td>270.61 198.41 157.20 229.78 405.19 346.68 320.78 297.38</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.07739 0.05484 0.03944 0.06334 0.10439 0.10115 0.10358 0.07581</td>
</tr>
<tr>
<td>Increase in MAD for new STARFM</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>0.01091 0.00609 -0.00181 -0.01116 -0.00112 0.00390 -0.00376 0.01640</td>
</tr>
<tr>
<td>NIR</td>
<td>0.02336 0.01240 -0.00404 0.00670 -0.00692 0.00389 -0.00515 0.01598</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.00001 0.00001 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000</td>
</tr>
<tr>
<td>STARFM v1.2.2</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>232.19 218.35 192.36 199.48 200.91 217.50 158.73 183.89</td>
</tr>
<tr>
<td>NIR</td>
<td>332.61 312.64 296.23 312.80 319.30 353.34 250.15 304.53</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.11673 0.10559 0.09511 0.08833 0.08767 0.06522 0.06025 0.09178</td>
</tr>
<tr>
<td>STARFM v1.1.2</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>232.17 218.36 192.37 199.48 200.91 217.51 158.73 183.88</td>
</tr>
<tr>
<td>NIR</td>
<td>332.58 312.62 296.23 312.80 319.29 353.34 250.17 304.52</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.11671 0.10559 0.09511 0.08833 0.08767 0.06522 0.06025 0.09177</td>
</tr>
<tr>
<td>Increase in MAD for new STARFM</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>0.01376 -0.00609 -0.00192 0.00250 -0.00041 -0.00651 0.00204 0.00853</td>
</tr>
<tr>
<td>NIR</td>
<td>0.02808 0.02012 -0.00074 0.00498 0.00497 -0.00293 -0.01747 0.01024</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.00002 0.00001 0.00000 0.00000 0.00000 0.00000 0.00000 0.00001</td>
</tr>
</tbody>
</table>
4. Discussion

4.1. Quality and Quantity of STARFM Predictions

The results show negligible difference in the quality of predictions made by each of the software versions. Although the MAD and $R^2$ values are different between the two versions, the difference is so small that when applied it should generally not make a difference. In addition, the newest version made the same amount of predictions as the original version. This indicates that the bug fix in v1.2.2 alleviated any concerns one may have when upgrading to the new version.

4.2. Limitations

Landsat scenes used to validate the predicted images were from the middle day of the 16-day composite window of the MODIS image. Therefore, the prediction could reflect any day within the 16-day composite window and thus a date different than the Landsat date used for validation of some or all of the pixels. This is a limitation in the ability to measure quality. If a perfect prediction was made and there were actual ground changes between the two dates of prediction and validation, the measured quality of the prediction should be reduced because of this.

In addition, the sample is limited to only eight clear days for one year produced by only two different base pairs. As reported, there is some variance between the predicted days and performance of base pairs. A larger set of validation dates and base pairs under various conditions may help provide a better understanding of the overall ability of STARFM.

Last, this study did not look at the distribution of fused locations. Although both software versions produced the same number of predictions, it was not determined if these are in the same
location. Visual inspection of a few predicted dates did not show any obvious differences in the locations of the predictions. However, it is possible that one software version produced more results in one area while the other version had equally more results in another area. Future work should examine the locations of predictions for biases.

5. Conclusions

Blending algorithms, such as the widely used STARFM, combine complementary remotely sensed data to better capture spatio-temporal dynamics (Emelyanova et al., 2013). STARFM’s originally released version of v1.1.2 was followed by v1.2.1, which improved computation efficiency (Gao et al., 2015). When Gervais et al. (2017) used both of these versions in an urban phenology study, generally v1.1.2 produced better results. However, those results did not conclude that the original version was superior. That study did not validate both versions while controlling for other variables in any robust way. Since that study a small bug was fixed in the newer v1.2.2 release. In addition, the previous study showed the new version produced fewer relative predictions toward the urban core. This left uncertainty around the decision to adopt the new version of the software or use the original version.

The current study addressed those above uncertainties by using both version 1.2.2 and the original STARFM version to fuse eight different dates produced from two different base pairs (16 total fused images). Results showed no difference in the number of predictions. In addition, the difference in the quality of the predictions between the two software versions was very small. The maximum difference in $R^2$ between the software versions for the red and NIR bands and calculated NDVI was less than 0.0001 and the maximum difference in MAD was less than 0.1 for both the scaled reflectance values of the red and NIR bands and less than 0.0001 for the
calculated NDVI values. Based on these results, and the improved computation efficiency of the new version of STARFM, it is reasonable to recommend adopting v1.2.2 in future studies, although it should be done with caution. This study did have few limitations, mainly because it tested the software under good conditions for a specific study area.
6. References


USDA-ARS. (n.d.). StarFM_ReadMe.
Chapter 4

Validating Urban Phenology from Fused Time Series Imagery with Citizen Science Data
for *Acer rubrum*

**Abstract:** Chapter 2 showed that when fused imagery are used to model the effects of the urban area on plant phenology, an earlier start of season (SOS), later end of season (EOS), and a longer length of season (LOS) are generally revealed in urban areas when compared to their surrounding rural counterparts. However, without in situ validation of those findings the results were uncertain. In particular, there remained the question of whether predicted phenophases are indicative of actual vegetative conditions, representative of other non-phenological events, such as land cover change, or simply the result of noise and errors. This paper helps to explain some of those uncertainties by comparing phenophase timings derived from fused images to in situ citizen science data of red maple (*Acer rubrum*). The results show that both datasets support the same general trends in phenology of earlier spring onset in urban settings. However, a statistically significant correlation between the two datasets for breaking leaf buds indicates red maple phenology is only representative of a fraction of the variance in phenology derived from fused images and it varies between cities.

1. Introduction

Phenological trends can be derived from both ground observations and remotely sensed data, each providing complementary information on different ecosystem traits. Ground observations describe individual plants (Badeck et al., 2004), and although useful, they are not spatially or temporally extensive. On the other hand, satellite measurements study phenology
over large areas (Hanes, Liang, & Morisette, 2014) but provide area-averaged information (Badeck et al., 2004). The so-called Land Surface Phenology (LSP) approach has documented the relationship between urban intensity and plant phenology (Walker, de Beurs, & Henebry, 2015). However, the detail in urban maps is dependent on the spatial resolution (Van de Voorde, Jacquet, & Canters, 2011) and the error in phenological detection increases as the temporal resolution decreases (Zhang, Friedl, & Schaaf, 2009). Image fusion may provide a means to improve analyses of phenology across urban-rural gradients by combining the relative strengths of different satellite sensors into a single product.

Liang et al. (2011) used a scaling approach to bring in situ observations and satellite pixels to a comparable scale and extent and successfully validated LSP, but they still encountered a number of limitations. The approach also involved a heavily supervised process step (Liang, Schwartz, & Fei, 2011). It was later applied to evaluate LSP extracted from fused images produced by the Spatial and Temporal Adaptive Reflectance Model (STARFM) for a heterogeneous mixed forest with fairly continuous canopy coverage. The results showed that, although the mean absolute error (MAE) of the fused LSP estimates at the community level was only about four days, the error range exceeded the difference between communities. They speculated that the fused product might be more useful in providing information in a more heterogeneous landscape (Liang et al., 2014).

While a number of blending, or image fusion, algorithms have been developed (for example Gao, Masek, Schwaller, & Hall, 2006, Hilker et al., 2009, Zhu, Chen, Gao, Chen, & Masek, 2010, D. Fu, Chen, Wang, Zhu, & Hilker, 2013, Huang & Song, 2012, and Gevaert & García-Haro, 2015) with the goal of combining different data sets to obtain more information than from each independent data source alone (Pohl & Van Genderen, 1998), STARFM has been
chosen by many studies because of its simplicity (Schmidt, Lucas, Bunting, Verbesselt, & Armston, 2015) and ability to accept one base pair (Liang et al., 2014). It was initially designed to predict surface reflectance data with Landsat spatial resolution (30m pixel) and MODIS temporal resolution (Gao et al., 2006) and has recently been applied to study differences in the start of season (SOS), end of season (EOS), and length of season (LOS) between urban and rural areas near Ogden, Utah. While earlier urban SOS and a later urban LOS were generally found, the lack of comparison with in situ observations leaves uncertainty around the results (Gervais, Buyantuev, & Gao, 2017).

The most obvious way to validate satellite-derived phenology is to use field observations. However, this approach is challenging because of the large area a pixel represents and the variability within it (Zhang, Friedl, & Schaaf, 2006). Scaling approaches, although successful, have shown drawbacks (Liang et al., 2011) and therefore may not be the best option in all situations. The aim of the present paper is to expand on the methods of Gervais et al. (2017) and investigate the applicability of using fused satellite data to assess the effects of the urban area on plant phenology. Specifically this paper will answer the following important research questions:

1. Are the results of phenology from fused imagery related to red maple (*Acer rubrum*) phenology from in situ observations of the USA National Phenology Network (USA-NPN)?

2. Do the results of phenology from fused imagery differ from in situ observations of the USA-NPN when observing the effects of the urban environment on red maple phenology?
To answer these questions, timings of spring and fall phenophases were extracted from a time series of Normalized Difference Vegetation Index (NDVI) values produced from fused images. These were then compared to phenophases observed with the in situ observations of the red maple USA-NPN data (https://www.usanpn.org/).

2. Methods

2.1. Study Area

Similar to previous studies, (i.e. Walker et al., 2015 and Krehbiel, Zhang, & Henebry, 2017), I delineated urban areas based on the US Census’s Cartographic Boundary Shapefile for Urban Areas (US Census Bureau, n.d.). To ensure thermal sums were the primary driver of phenology, I limited my dataset to locations east of 97°W and north of 35°N similar to Liang and Schwartz (2014). Based on the previous LSP study by Walker et al. (2015), my analysis of rural areas was restricted to 40km beyond the urban boundary.

2.2. In Situ Data

The primary source for ground observations was citizen science data. Citizen science is described as “the practice of engaging the public in a scientific project – a project that produces reliable data and information usable by scientists, decisionmakers, or the public and that is open to the same system of peer review that applies to conventional science” (McKinley et al., 2015, p. 3). Citizen science does not only engage the public, but also allows for science to be done that may not otherwise be feasible due to scale or other reasons, with a major strength in its ability to collect fine-grained information over large areas and time (McKinley et al., 2015).
Participation by the general public will be a necessary component of successful phenological monitoring activities. The contemporary USA-NPN was established in 2007 to promote an understanding of phenology, by making phenology data freely available and also providing standardized protocols for phenology monitoring which enable individuals and groups to collect and organize phenology observations at the national scale (Schwartz, Betancourt, & Weltzin, 2012). The monitoring methods assess the presence or absence of a phenophase which can then be used to determine the date of an event (Denny et al., 2014).

In addition to the quality assurance and quality control processes employed by the USA-NPN (see https://www.usanpn.org/data/quality), I implemented two additional quality control criteria on the dataset. First, similar to Gerst, Rossington, and Mazer (2017) and Elmore, Stylinsky, and Pradhan (2016), I ensured timeliness and temporal accuracy of observations by including only positive observations that were within a certain number of days of a negative observation, which is indicated by the “NumDays_Since_Prior_No” attribute of the database. The seven day threshold I selected was the more restrictive of the two, the same as Gerst et al. (2017). Second, any observations with a positive “Observed_Status_Conflict_Flag” were removed.

2.3. City and Species Selection

Individual phenometrics data for the dates that had Landsat 8 data available (at the time of this study) over the entire expected temporal range for any given phenophase were downloaded from the USA-NPN Phenology Observation Portal (http://data.usanpn.org/observations/get-started, downloaded on Jan 17, 2018). The resulting data showed that the red maple is the most commonly observed plant species in the study area within
the United States between day of year (DOY) 200 2013 – DOY 365 2017. It is a late-successional species that is considered a “supergeneralist” due to its ability to grow in the widest variety of sites and conditions (United States Department of Agriculture, Natural Resource Conservation Service, n.d.-a). It has a wetland classification of FAC (United States Department of Agriculture, Natural Resource Conservation Service, n.d.-b) indicating that it can occur in either wetlands or non-wetlands (United States Department of Agriculture, Natural Resource Conservation Service, n.d.-c).

I selected cities that had the largest number of comparable quality urban and exurban observations for the red maple with seven day accuracy and no observer conflict, in order to have the largest sample size with the fewest sets of imagery to process, thus maintaining high efficiency. In this context, an exurban observation is considered to be comparable to an urban observation if there was an urban observation for the same city, phenophase, and year, and vice versa. In addition, any cities that had urban or exurban observations near (i.e. within 40km of) the Atlantic coast, any bay of the Atlantic, or the Great Lakes, were excluded. The cities of Asheville, NC, Blacksburg, VA, and Minneapolis–St. Paul, MN—WI, are the three with the largest number (>85) of comparable quality observations and were selected for analysis (Figure 4-1).
2.4. USA-NPN Phenophase Selection and Quality Control

The phenophase category of leaves had the greatest number of comparable observations for the three selected cities and was used for analysis. There are five phenophase descriptions in this category that are monitored in the USA-NPN: breaking leaf buds, leaves, increasing leaf size, colored leaves, and falling leaves. For these five phenophase descriptions, referred to hereafter as simply phenophases, I implemented three additional quality control measures on the USA-NPN data. First, only spring phenophase observations within a date range were used to ensure the earliest onset date of each individual is included and to identify outliers (i.e. Elmore et al., 2016, Gerst et al., 2017, Piao et al., 2015, Liang & Schwartz, 2014, Y. H. Fu et al., 2015). The range I set for the spring applied to the breaking leaf buds, leaves, and increasing leaf size phenophases. Being the least restrictive range used in previous studies it was limited to observations recorded before DOY 200, which was the maximum upper threshold used by Elmore et al. (2016) for all species. Second, in a similar manner, fall leaf phenophases (colored
leaves and falling leaves) were limited to observations for the second part of the year, DOY 200-365, once again the same as Elmore et al. (2016). Third, one urban observation location in MN was found to be in a lake and one exurban observation location in NC was frequently (75% of the time) flagged as unclear in potential base pairs (see next section for definition of base pairs). All observations from these two locations were removed from analysis as well as any corresponding exurban points that had no other corresponding urban points. The frequencies of comparable phenophase observations that meet these criteria for the three cities are presented in Table 4-1.

**Table 4-1.** Frequency of phenophases with comparable observations for the red maple for the selected study cities.

<table>
<thead>
<tr>
<th>Phenophase/City</th>
<th>Urban Observation</th>
<th>Exurban Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breaking leaf buds (all)</td>
<td>20</td>
<td>89</td>
</tr>
<tr>
<td>NC</td>
<td>11</td>
<td>72</td>
</tr>
<tr>
<td>VA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MN</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>Falling leaves (all)</td>
<td>38</td>
<td>51</td>
</tr>
<tr>
<td>NC</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>VA</td>
<td>28</td>
<td>13</td>
</tr>
<tr>
<td>MN</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Colored leaves (all)</td>
<td>33</td>
<td>49</td>
</tr>
<tr>
<td>NC</td>
<td>6</td>
<td>40</td>
</tr>
<tr>
<td>VA</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>MN</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Increasing leaf size (all)</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>NC</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>VA</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>MN</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Leaves (all)</td>
<td>28</td>
<td>26</td>
</tr>
<tr>
<td>NC</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>VA</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>MN</td>
<td>14</td>
<td>16</td>
</tr>
</tbody>
</table>
2.5. Image Fusion

STARFM was developed to combine high spatial resolution satellite data, i.e. Landsat, with the temporal frequency of a coarse resolution sensor, i.e. the Moderate Resolution Imaging Spectroradiometer (MODIS) (Gao et al., 2006). It is based on the assumption that temporal changes observed in MODIS imagery indicate expected changes at the Landsat scale (Gao et al., 2015). Although it was originally used to predict daily surface reflectance at the Landsat spatial resolution by fusing Landsat-7 ETM+ images and 500m MODIS daily surface reflectance, the approach is also applicable to the Landsat OLI (Operational Land Imager) sensor (Gao et al., 2006). Since it was first developed, several MODIS data products have been used with STARFM, with the MODIS Nadir Bidirectional Reflectance Distribution Function (BRDF) Adjusted Reflectance (NBAR) 16-day composite data (MCD43A4) performing the best out of the three that were compared (Walker, de Beurs, Wynne, & Gao, 2012). The selected data used in the fusion for this study included Landsat 8 OLI and MODIS MCD43A4 V006 data, the same as in Chapter 3. The study areas span seven Landsat scenes (Path/row 27/28, 27/29, 18/35, 18/36, 28/28, 19/35, and 17/34), however scenes 28/28 and 19/35 only offer redundant spatial coverage of the study areas and were not processed. All study areas are completely contained in the two MODIS scenes of h11v04 and h11v05, data for both of which were downloaded and used.

Image base pairs were selected, data preprocessed, and STARFM parameters set according to the documentation in Gervais et al. (2017) for the statistical selection criteria and the three minor modifications previously explained in Chapter 3 including the use of Pixel_QA values of 322 and 386 as clear. For the purpose of base pair selections, the area analyzed for clarity was defined as the area around each point which was within the search distance of STARFM, or 1500m as in Gervais et al. (2017). For MN and VA, a Landsat image was used as a
base pair if this area was at least 95% clear for all areas that are within its extent. For NC, this was relaxed to 90%. However, in 2014 no images met this criteria for NC. To compensate for this, the Landsat image that was most clear (>88%) for all areas that correspond to 2014 observations was used. In addition, all Landsat scenes from the same date for NC and MN were mosaicked before processing because these study areas each span more than one Landsat scene. One additional quality assurance (QA) step was also implemented to check and ensure the focal pixel of the USA-NPN observation was clear for the base pairs. The only exception to this was year 2014 in NC, where an observation location from another year had an unclear focal pixel. The MODIS value was substituted into the fine resolution image in this situation. The selected base pairs are shown in Figure 4-2.

After preprocessing, STARFM v1.2.2 was run on Ubuntu operating system (https://www.ubuntu.com/) to produce fused images every five days for the date range of expected seasonal dates that coincide with USA-NPN observations. Since 2016 was a leap year, DOY 366 was omitted. This allowed for the same number of predictions and for the same DOY of predictions every year. Although MN did not have observations in 2014 and VA did not have any in 2015, images were still fused in these areas for this year since this year is encompassed by other years with observations. There were also no observations for VA and NC in 2013, and since this was the first year there were no images fused for this year for these areas.

To ensure base pairs were able to produce acceptable fused images, a simple validation of the NDVI values derived from fused images was performed, which consisted of comparing, via mean absolute difference (MAD) and $R^2$ similar to Walker, de Beurs, Wynne & Gao (2012) and Gervais et al. (2017), for example, the NDVI values of focal pixels of the USA-NPN observations to actual Landsat values of the same date. Because the validation date may not
coincide with the five day sample time, I fused additional images for validation that may not have been used in the final analysis. The validation dates chosen had many clear focal pixels for the base pairs. However, the base pair for VA from 2017 DOY 144 was not validated because there were two other base pairs in close temporal proximity.

![Graph showing DOY of Base Pairs by Year](image)

**Figure 4-2.** Base pairs used in the image fusion.

2.6. Extraction of Phenophase Event Dates

Before any seasonality parameters could be extracted, a time series of vegetation indices was prepared from the fused images produced by STARFM. The vegetation index used was the Normalized Difference Vegetation Index (NDVI). This index allows an estimation of seasonal changes that are the result of phenological switches (Badeck et al., 2004).

TIMESAT, implemented in a computer program, was used to extract the time of the SOS and EOS event from the NDVI time series that was produced from fused imagery. It smooths a time-series of satellite data, such as NDVI, to accurately extract phenological parameters.
(Jönsson & Eklundh, 2002). The SOS is defined as a point when the value has increased by a
certain percent and the EOS in a similar way (Jönsson & Eklundh, 2004). The value selected was
20%, which is the default amplitude (Tan et al., 2011) and the same as Gervais et al. (2017). All
other parameters were set to the same as in this previous study. Since TIMESAT only extracts
seasonality for the n-1 center years (Eklundh & Jönsson, 2015), data for a year before and a year
after were created and used but not interpreted.

2.7. Comparing Phenophases

Similar to Elmore et al. (2016), a focal pixel for the phenology data derived from the
fused images was identified from the USA-NPN observations. To make a direct comparison
between the phenology data from the fused images and the USA-NPN data, values for these
focal pixels were selected for analysis. The dates of SOS were compared to the locations of each
of the three spring phenophases and the dates of EOS were compared to the locations with fall
phenophases. To see how well variation in the predicted dates of EOS and SOS are related to the
variation in the observed dates of the USA-NPN data, $R^2$ values were calculated.

As described in Gervais et al. (2017), prior to calculating the urban effect on phenology
the USA-NPN data for each city were normalized. The mean value for each phenophase of the
exurban area of each city and year was subtracted from each individual value of the same city,
phenophase, and year. Phenology derived from fused data was normalized in a similar way for
each year and city. To ensure the same locations were used for normalization of phenology
derived from the fused data as were used to normalize USA-NPN phenology, phenology from
the fused data was normalized according to the USA-NPN phenophase of the focal pixel and not
simply by EOS or SOS for each city and year. The normalization allowed for comparison
between years (Gervais et al., 2017). After normalization, MAD between each point, average urban advancement, and $R^2$ were produced for each phenophase and city, allowing for a comparison of the SOS and EOS calculated from the fused images with the five phenophases extracted from the USA-NPN data.

3. Results

3.1. Image Fusion Validation

The average $R^2$ values for all days is lowest (0.53) in the MN study area. Both NC and VA have an average $R^2$ values greater than 0.8 for all days. The MAD followed a similar trend, with both VA and NC having a MAD of 0.06 and MN having a MAD of 0.07 (Table 4-2).
Table 4-2. Validation results by DOY for each of the cities and exurban areas.
(a) Minneapolis/St. Paul, MN-WI; (b) Asheville, NC; (c) Blacksburg, VA.

(a) Minneapolis/St. Paul, MN-WI

<table>
<thead>
<tr>
<th>Year</th>
<th>DOY</th>
<th>Average $R^2$</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>243</td>
<td>0.94</td>
<td>0.037</td>
</tr>
<tr>
<td>2014</td>
<td>150</td>
<td>0.41</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>198</td>
<td>0.82</td>
<td>0.049</td>
</tr>
<tr>
<td>2015</td>
<td>73</td>
<td>0.28</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>313</td>
<td>0.24</td>
<td>0.095</td>
</tr>
<tr>
<td>2016</td>
<td>60</td>
<td>0.35</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>1.00</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>316</td>
<td>0.34</td>
<td>0.100</td>
</tr>
<tr>
<td>2017</td>
<td>46</td>
<td>0.66</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>158</td>
<td>0.70</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>334</td>
<td>0.12</td>
<td>0.126</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.53</td>
<td>0.07</td>
</tr>
</tbody>
</table>

(b) Asheville, NC

<table>
<thead>
<tr>
<th>Year</th>
<th>DOY</th>
<th>Average $R^2$</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>103</td>
<td>0.80</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>263</td>
<td>0.89</td>
<td>0.058</td>
</tr>
<tr>
<td>2015</td>
<td>122</td>
<td>0.83</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.66</td>
<td>0.076</td>
</tr>
<tr>
<td>2016</td>
<td>269</td>
<td>0.89</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>317</td>
<td>0.83</td>
<td>0.044</td>
</tr>
<tr>
<td>2017</td>
<td>127</td>
<td>0.88</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>271</td>
<td>0.93</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>303</td>
<td>0.90</td>
<td>0.054</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.85</td>
<td>0.06</td>
</tr>
</tbody>
</table>

(c) Blacksburg, VA

<table>
<thead>
<tr>
<th>Year</th>
<th>DOY</th>
<th>Average $R^2$</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>80</td>
<td>0.84</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>192</td>
<td>0.86</td>
<td>0.058</td>
</tr>
<tr>
<td>2015</td>
<td>35</td>
<td>0.81</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>83</td>
<td>0.62</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>259</td>
<td>0.78</td>
<td>0.080</td>
</tr>
<tr>
<td>2016</td>
<td>86</td>
<td>0.82</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>134</td>
<td>0.82</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>0.92</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>278</td>
<td>0.62</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>342</td>
<td>0.94</td>
<td>0.051</td>
</tr>
<tr>
<td>2017</td>
<td>72</td>
<td>0.93</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>0.88</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.88</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>328</td>
<td>0.90</td>
<td>0.049</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.83</td>
<td>0.06</td>
</tr>
</tbody>
</table>
3.2. Variability of Phenophases

For all cities together, the relationship (expressed as $R^2$) between the DOY of phenology from the fused imagery and USA-NPN phenophases is below 0.04 for all phenophases with breaking leaf buds and falling leaves being the only statistically significant phenophases at the 90% confidence level. The relationship between the normalized values is highest for the breaking leaf bud phenophase locations and insignificant in the fall. Both the breaking leaf buds and leaves are statistically significant for the normalized values at the 99.9% and 99% levels, respectively (Table 4-3).

When looking at the cities separately, more times than not the $R^2$ values for both the DOY and normalized data increase when compared to the values for all cities together. For spring phenophases, the breaking leaf buds has the highest statistical significance for each of the cities with observations. The $R^2$ values for this phenophase of the DOY data range from 0.41 to 0.46 and $R^2$ values of the normalized data range from 0.34 to 0.51. There is not as clear of a pattern for the fall phenophases, with only the colored leaves of the DOY data for VA and the normalized values for MN being statistically significant at the 95% level (Table 4-3).
Table 4-3. $R^2$ values between phenophases of the USA-NPN and phenology derived from fused imagery for both DOY and normalized data. Significant values are in bold and significance level is indicated by: *** 0.001, ** 0.01, * 0.05, . 0.1.

<table>
<thead>
<tr>
<th>Phenophase</th>
<th>Location</th>
<th>$R^2$ DOY All Areas</th>
<th>$R^2$ Normalized Values All Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breaking leaf buds</td>
<td>All cities</td>
<td>0.03 .</td>
<td>0.33 ***</td>
</tr>
<tr>
<td></td>
<td>NC</td>
<td>0.41***</td>
<td>0.51 ***</td>
</tr>
<tr>
<td></td>
<td>MN</td>
<td>0.46 ***</td>
<td>0.34 **</td>
</tr>
<tr>
<td>Increasing leaf size</td>
<td>All cities</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>NC</td>
<td>0.46**</td>
<td>0.46**</td>
</tr>
<tr>
<td></td>
<td>VA</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>MN</td>
<td>1.4E-05</td>
<td>0.03</td>
</tr>
<tr>
<td>Leaves</td>
<td>All cities</td>
<td>0.01</td>
<td>0.14**</td>
</tr>
<tr>
<td></td>
<td>NC</td>
<td>0.30*</td>
<td>0.30*</td>
</tr>
<tr>
<td></td>
<td>VA</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>MN</td>
<td>0.27**</td>
<td>0.16*</td>
</tr>
<tr>
<td>Colored leaves</td>
<td>All cities</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>NC</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>VA</td>
<td>0.16*</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>MN</td>
<td>0.21</td>
<td>0.44*</td>
</tr>
<tr>
<td>Falling leaves</td>
<td>All cities</td>
<td>0.03 .</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>NC</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>VA</td>
<td>0.06</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>MN</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

3.3. Average Advancement of Seasons

Averaged results for the normalized USA-NPN data and phenology from the fused imagery show some consistency between phenophases in relation to the urban phenology across the aggregate of all cities. All of the spring phenophases occur earlier in the aggregated urban areas. This trend is also apparent when looking at each city individually, with all except one phenophase in one city for the USA-NPN data having urban spring phenophases earlier. Autumn
phenophases show inconsistent advancement for the USA-NPN data across phenophases and cities with half of the aggregate locations and some of the city level data showing later urban fall phenophases. The urban EOS advancement from the fused data are negative (earlier) for the two autumn phenophase locations of all areas, but the advancement is less than a day. The EOS from fused data are also inconsistent between cities when looking at the city level results for fall phenophases with NC showing a delayed urban fall and the other cities showing an earlier urban fall for both phenophases (Table 4-4).

The MAD for individual normalized points and the difference between the average urban advancement from the USA-NPN and phenology from fused data in the aggregate of the cities is smallest for the spring phenophase of breaking leaf buds and for the fall phenophase colored leaves. For spring phenophases, in MN the difference between the average urban advancement of the USA-NPN and SOS from fused data is smallest (8.1 days) for the leaves phenophase but the MAD of individual points is smallest (7.9 days) for the breaking leaf buds phenophases. NC shows a smallest difference for the averages of only 4.1 days for increasing leaf size and the smallest MAD of 8.8 days for leaves. As noted previously, there is no data for VA breaking leaf buds and the two spring phenophases for VA show an average difference in urban advancement and MAD of at least twice as large as the other two cities. Fall phenophases are more consistent within each city, with VA and MN both showing the colored leaves phenophase having the smallest values for the two measurements and NC shows falling leaves having the smallest values (Table 4-4).
Table 4-4. Average urban advancement (negative) or delay (positive) for phenophases based on the USA-NPN and fused data. The difference between the two averages for each phenophase and Mean Absolute Difference (MAD) in phenology for individual locations is also reported.

<table>
<thead>
<tr>
<th>Phenophase</th>
<th>Location</th>
<th>Days earlier (negative) urban vs exurban USA-NPN data</th>
<th>Days earlier (negative) urban vs exurban Fused data</th>
<th>Absolute Difference Between Average USA-NPN and Fused Advancement</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breaking leaf buds</td>
<td>All cities</td>
<td>-25.6</td>
<td>-17.6</td>
<td>8.0</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>Asheville, NC</td>
<td>-39.7</td>
<td>-16.6</td>
<td>23.0</td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>Minneapolis--St. Paul, MN--WI</td>
<td>-8.3</td>
<td>-18.7</td>
<td>10.4</td>
<td>7.9</td>
</tr>
<tr>
<td>Increasing leaf size</td>
<td>All cities</td>
<td>-4.1</td>
<td>-27.7</td>
<td>23.6</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td>Asheville, NC</td>
<td>-19.4</td>
<td>-15.3</td>
<td>4.1</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td>Blacksburg, VA</td>
<td>-11.6</td>
<td>-61.1</td>
<td>49.5</td>
<td>44.0</td>
</tr>
<tr>
<td></td>
<td>Minneapolis--St. Paul, MN--WI</td>
<td>5.4</td>
<td>-15.7</td>
<td>21.1</td>
<td>16.0</td>
</tr>
<tr>
<td>Leaves</td>
<td>All cities</td>
<td>-7.5</td>
<td>-22.0</td>
<td>14.5</td>
<td>17.3</td>
</tr>
<tr>
<td></td>
<td>Asheville, NC</td>
<td>-0.9</td>
<td>-11.1</td>
<td>10.2</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Blacksburg, VA</td>
<td>-10.8</td>
<td>-37.4</td>
<td>26.6</td>
<td>51.5</td>
</tr>
<tr>
<td></td>
<td>Minneapolis--St. Paul, MN--WI</td>
<td>-7.9</td>
<td>-15.9</td>
<td>8.1</td>
<td>9.9</td>
</tr>
<tr>
<td>Colored leaves</td>
<td>All cities</td>
<td>2.0</td>
<td>-0.2</td>
<td>2.2</td>
<td>16.1</td>
</tr>
<tr>
<td></td>
<td>Asheville, NC</td>
<td>17.1</td>
<td>4.5</td>
<td>12.7</td>
<td>20.3</td>
</tr>
<tr>
<td></td>
<td>Blacksburg, VA</td>
<td>-1.4</td>
<td>-1.5</td>
<td>0.1</td>
<td>12.7</td>
</tr>
<tr>
<td></td>
<td>Minneapolis--St. Paul, MN--WI</td>
<td>-1.4</td>
<td>-0.3</td>
<td>1.1</td>
<td>6.7</td>
</tr>
<tr>
<td>Falling leaves</td>
<td>All cities</td>
<td>-5.5</td>
<td>-0.8</td>
<td>4.7</td>
<td>21.2</td>
</tr>
<tr>
<td></td>
<td>Asheville, NC</td>
<td>3.9</td>
<td>4.9</td>
<td>1.0</td>
<td>19.4</td>
</tr>
<tr>
<td></td>
<td>Blacksburg, VA</td>
<td>-12.6</td>
<td>-0.6</td>
<td>12.0</td>
<td>22.0</td>
</tr>
<tr>
<td></td>
<td>Minneapolis--St. Paul, MN--WI</td>
<td>30.1</td>
<td>-10.9</td>
<td>41.0</td>
<td>26.5</td>
</tr>
</tbody>
</table>
4. Discussion

4.1. Accuracy Assessment of STARFM

Overall, the accuracy assessment of STARFM’s ability to predict a Landsat NDVI image is better for both NC and VA than for MN. Compared to the average $R^2$ values of slightly above 0.6 in Chapter 3, the MN results tend to be slightly less accurate than expected but the NC and VA results are better than expected. However, compared to the average MAD results of approximately 0.8 (Chapter 3), all areas performed slightly better overall.

4.2. Variability

The relationship between USA-NPN breaking leaf buds and the spring phenology from fused data is the only consistently significant relationship, and it also has the highest normalized $R^2$ values for the cities that had observations for this phenophase. However, the $R^2$ value for the normalized data ranged from only 0.34 in MN to 0.51 in NC. This suggests that for each of the cities at best only about half of the variation in the timing of SOS across years, relative to the mean exurban phenophase timing, derived from fused data can be explained by the variation in the relative timing of the red maple breaking leaf buds as observed by citizen scientists. However, the results from fused imagery could be more closely related to actual phenophase timing than reported here. The NDVI signal from the fused imagery may actually represent phenology of a species other than the red maple (i.e. another dominant species in the pixel) or a combination of species where the red maple is only contributing to a fraction of the area (i.e. the mixed pixel problem). For example, in a heterogeneous environment, understory plants will unfold their leaves before trees and will advance satellite measurements (Badeck et al., 2004), so the phenology from the fused imagery may represent variation in phenophases of an understory
The $R^2$ values between aggregated DOY data from all cities for the USA-NPN and phenology derived from fused data are fairly low, which means that the variability in the phenophases of the red maple from year to year and city to city is not explained well by the phenology derived from the fused data. However, the $R^2$ values generally increased when each city was considered individually, which indicates that the relationship between phenophases of the red maple and the fused data is not the same for each city. This could indicate that the species composition of a mixed pixel is varying across locations and the response to the urban area differed depending on the species. Alternatively, it could indicate that the response of the red maple to the urban area varies across locations.

The general increase in the spring relationship after normalization for the aggregated cities suggests that some inter-city variation may be accounted for in the normalization process. However, when looking at the cities separately, normalization increases the $R^2$ values as many times as it decreases them and slightly lowers the significance for three phenophases and only increases it for one. This suggests that this method of normalization may inconsistently account for variation in phenology at the city level.

4.3. Average Normalized Values

Since leaves appear earlier in warmer springs (Wesołowski & Rowiński, 2006) and temperatures are higher in urban environments, referred to as the urban heat island (UHI) (Oke, 1982), it makes sense that almost all locations show an earlier spring on average for both USA-NPN data and phenology from fused data. For the USA-NPN data, for all cities combined as well
as MN and NC, breaking leaf buds has the largest urban advancement when compared to exurban areas. Advancement in spring phenology for individual cities from the fused data for NC and MN shows an advancement in urban areas between about 11 to 17 days and 16 to 19 days for the two. For the phenology from the fused dataset, these differences in timing between the locations of phenophases within a city were not expected since the same amplitude of change is assessed, whereas the USA-NPN data assessed different ground events. However, the range in timing of phenophase dates between locations is generally small for both SOS and EOS based on the fused data compared to the USA-NPN data (i.e. 5.5 days versus 38.8 days for spring in NC).

The exception to the range of about 11 to 19 days for spring urban advancement from fused data is VA, which shows an average urban advancement of over 61 days for increasing leaf size locations, 37 days for leaves locations, and does not have any data for breaking leaf buds. It is possible that the data in these VA locations contain errors, leading to an erroneous increase in the average advancement for the aggregate of cities as well. The USA-NPN data does not show a similar pattern for the VA locations. However, since there is no evidence besides the number itself being large, the data was included in the aggregate level analysis and could be accurate.

The difference between the USA-NPN and fused based average advancement in the spring for the aggregate cities is the smallest for the breaking leaf buds phenophase. Interestingly, none of the cities alone show the smallest difference of the averages for breaking leaf buds even though it had the strongest relationship in NC and MN. In addition, the MAD for the points is the lowest for varying phenophases across the cities and does not consistently coincide with the highest or most statistically significant $R^2$ value.

The largest average urban advancement of fall phenophases for the aggregate cities is much smaller than the largest advancement of spring for the aggregate cities for the phenology
from both the fused data and USA-NPN data. The colored leaves phenophase is even delayed for all urban areas combined according to the USA-NPN data. This generally small difference may be caused by photoperiod acting as the main cue for growth cessation, which is consistent along the urban to rural gradient.

4.4. Limitations and Future Work

Although citizen science data of the USA-NPN have been shown to be accurate (Fuccillo, Crimmins, de Rivera, & Elder, 2015), they still contain errors. As Fuccillo et al. (2015) explained, it is also reasonable to assume that the errors are not uniform throughout all observations, possibly because some individuals have more and some have less experience (Fuccillo et al., 2015). Although some errors were removed before data processing with the QA/QC steps taken by the USA-NPN (see https://www.usanpn.org/data/quality) and then the additional steps taken in this study, there may remain other less obvious errors in the data. Since there is no way of knowing if the observations used here contain errors, a larger sample size may help in building more confidence. This study used the largest possible sample size while maintaining a manageable amount of data to process, but the resulting subset may still be smaller than the desired size. With more data being collected each year, future research can expand on the ideas presented here and include a larger sample size that may be more representative of the real-world conditions in their entirety.

I did not attempt to co-register Landsat and MODIS pixels used in image fusion. As Gao et al. (2015) explained, MODIS data have a good geolocation accuracy at the 500m resolution, but the level of accuracy is not enough for the Landsat spatial resolution at which the fusion is
performed (Gao et al., 2015). Future work could attempt to improve the fusion results further by first co-registering MODIS data to Landsat data.

Some of the ranges of the phenology from the fused data appear unexpected. Several of the EOS go well into the following year and one SOS is found in August. Applying a QA/QC process similar to the ones implemented on the USA-NPN data may be appropriate to remove this data if further evidence (i.e. NDVI values or fitted curves) appear erroneous.

This study specifically focused on the leaves phenophases of the red maple. However, other previous studies have shown that species besides the red maple can be more correlated with (Elmore et al., 2016) and have their buds break closer to the predicted date of satellite based phenology (Parece & Campbell, 2018). In addition, lilac (*Syringa × chinensis*) full flowering has been shown to be correlated with satellite based phenology better than leaves and breaking leaf buds phenophases (Elmore et al., 2016). Future work should not limit validation of phenology derived from fused imagery to just the leaves phenophases or the red maple and explore other possibilities.

Last, TIMESAT offers a variety of settings that are used to determine the values of seasonality parameters. This study did not explore the effects of changing values for the settings, which could impact the results. For example, previous research has shown some intra-urban variability of predicted phenology based on fitting method (Parece & Campbell, 2018). In addition, different parameter setting in STARFM were not tested. Future research could conduct a sensitivity analysis on changing the setting values to see if other options produce better results.
5. Conclusion

This study compared the leaves phenophases of USA-NPN red maple data to phenophase dates derived from fitted curves of NDVI values, which were calculated from fused remotely sensed imagery using STARFM. Both methods provided evidence of expected spring and fall phenophase advancement in urban areas, showing an earlier arrival of spring in urban areas and less pronounced difference in the fall. The advancement of breaking leaf buds is overall the closest between the phenology from fused imagery and USA-NPN data, being only eight days different with some variation between cities. For fall, the colored leaves phenophase advancement of USA-NPN data is closest to the fused based phenology advancement with the difference of 2.2 days, also with inter-city variation.

There is uncertainty around these results since no single USA-NPN phenophase is consistently closest to the fused based phenophases between the cities. In addition, the amount of variation in the phenology from fused data that is explained by red maple USA-NPN phenophases is generally less than half, although statistically significant. Such low level of the relationship with the USA-NPN red maple phenophases may mean that the methods used here to estimate phenophases from remotely sensed data are either only partially represented by red maple phenology or at least one of the datasets lacks accuracy. However, the average urban phenophase advancement revealed by the fused data does appear reasonable and may correlate better to another species, combination of species, or phenophase. Future work should attempt to validate these methods for a variety of species and phenophases besides those of red maple leaves.
6. References:


1. General Conclusion

Fusing remotely sensed images together with the Spatial and Temporal Adaptive Reflectance Model (STARFM, see Gao, Masek, Schwaller, & Hall, 2006) to analyze phenology across urban to rural gradients offers possible advantages over using in situ observations or remotely sensed data from a single satellite alone. In this dissertation, in Chapters 2-4, I was able to produce and validate fused images from successive software versions of STARFM and across the urban to rural gradient of four cities of the United States. The resulting fused images had a spatial resolution of Landsat and a temporal resolution better than Landsat. Based on previous literature, spatial resolution affects the detail of urban maps (Van de Voorde, Jacquet, & Canters, 2011) while temporal resolution is important for phenology detection (Zhang, Friedl, & Schaaf, 2009). Therefore, the traits of the fused data should improve the ability to monitor urban phenology over large heterogeneous areas. I was able to document evidence of the expected variation in urban phenology with the results from fused data between land cover classes (Chapter 2). Chapter 3 examined the differences between the original version and a new version of STARFM and demonstrated that the new version should be used in future work. However, as shown in Chapter 4, when comparing predicted phenophases from the fused data with in situ citizen science data for the red maple \textit{(Acer rubrum)} leaves phenophases (made available by the USA National Phenology Network (USA-NPN)), at most only 51% of the variance in the phenology from the fused data relative to the mean exurban phenophase timing was explained by the in situ data, with relationships varying between cities. It is likely that more than just the red
maple is contributing to the fused phenophase dates, which urges for the evaluation of other species and phenophases in the future.

This topic of monitoring effects on phenology in the urban area is important. Urban areas are generally warmer than the surrounding rural environment, a phenomenon referred to as the Urban Heat Island (UHI) (Oke, 1982), and may offer a glimpse into effects of future climate change (i.e. Ziska et al., 2003). The evidence presented in Chapters 2 and 4 supports the previously reported results that plants generally bloom earlier in warm springs (Wesołowski & Rowiński, 2006). My results from fusing satellite data to extract phenophases show an average urban spring advancement from 0.28 to 61.1 days, depending on the location. The results from citizen science data for the red maple also show an urban spring advancement in each city for each phenophases ranging from 0.9 to 39.7 days except for the increasing leaf size in Minneapolis which showed an urban delay of 5.4 days. However, it has been previously shown that photoperiod drives fall events (Perry, 1971). Once again, my results support this observation since the effects of the urban area on fall phenophases is not as clear or consistent. The results from citizen science data show an urban fall advancement from 12.6 days up to a fall urban delay of 30.1 days, depending on the city and phenophase. The results from fused data show an urban advancement of 10.9 days to an urban delay of 4.5 days.

2. Existing Gaps in Knowledge and Future Work

Several unanswered questions and uncertainties were left outside the scope of this research. First, there are a number of fusion methods available but this research focused on only one, STARFM. Future work should examine the relationship between red maple phenology and phenology derived from other fusion methods. Second, only one set of parameters was used to
produce and extract phenophases from the fused data. Testing other parameters may yield different results. Third, the USA-NPN data used had several limitations. In at least one situation there was only one exurban value used to normalize the urban points. If this one point was incorrect, the urban effect would also be incorrect for all urban points associated with it. A larger dataset would help identify outliers and reduce their effects. At the present time, these data are limited, but every day more observations are provided to the USA-NPN.
3. References:


