

Article

Detection of Emerging Stress in Trees Using Hyperspectral Indices as Classification Features

Laura M. Moley ¹, Douglas G. Goodin ^{1,*} and William P. Winslow III ^{1,2}

¹ Department of Geography and Geospatial Sciences, Kansas State University, Manhattan, KS 66503, USA; Imoley@ksu.edu (L.M.M.); chipwin@tamu.edu (W.P.W.III)

² Department of Landscape Architecture and Urban Planning, Texas A&M University, College Station, TX 77843, USA

* Correspondence: dgoodin@ksu.edu

Abstract: This research presents a classification methodology for the detection of new or emerging stress in trees using indices derived from hyperspectral data and tests whether existing hyperspectral indices are effective when used as the classification features for this problem. We tested six existing indices—Water Band Index (WBI), Gitelson–Merzlyak B Index (GMB), Normalized Phaeophytization Index (NPQI), Combined Carotenoid/Chlorophyll Ratio Index (CCRI), Photochemical Reflectance Index (PRI), and Red-Edge Chlorophyll Index (Cire)—along with a seventh Test Index—generated as a composite of PRI and Cire—as classification features. Analysis was conducted using data collected from trees with and without emerald ash borer (EAB) infestation to develop a methodology that could be adapted to measure emerging stress from other pathogens or invasive pests in other tree species. Previous work has focused specifically on the identification of damage or stress symptoms caused by a specific pathogen. In this study, we adapted that work to develop a system of classification that can be applied to the identification of stress symptoms from a range of sources, measurable in trees based on spectral response and, in some cases, detectable prior to the onset of visible symptoms that can be measured through human observation. Our data indicate that existing indices derived from hyperspectral data are effective as classification features when measuring spectral responses indicative of emerging stress in trees.

Keywords: tree pathology; urban forestry; environmental stress; hyperspectral indices; vegetation classification; suburban landscapes; disease spread; remote sensing



Citation: Moley, L.M.; Goodin, D.G.; Winslow, W.P., III. Detection of Emerging Stress in Trees Using Hyperspectral Indices as Classification Features. *Environments* **2024**, *11*, 85. <https://doi.org/10.3390/environments11040085>

Academic Editor: Sergio Ulgiati

Received: 1 March 2024

Revised: 8 April 2024

Accepted: 16 April 2024

Published: 22 April 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

One major challenge in plant research is the development of more sensitive techniques for the early detection of plant stress. Remote sensing techniques, particularly at the leaf level, have proven to be valuable tools for collecting such data. In particular, hyperspectral remote sensing has proven to be a promising technique for detecting stress response in a variety of plant types, including trees and woody vegetation [1,2]. Hyperspectral remote sensing, which can be broadly defined as the acquisition of reflectance in the electromagnetic spectrum in a large number of contiguous spectral bands [3], is effective in remote sensing studies of vegetation stress because it resolves the response from surfaces into much finer spectral detail, allowing for the detection of reflectance and absorption features which might not be detectable to multispectral remote sensing systems with coarser spectral response [4]. This is particularly useful in the analysis of vegetation, where physiological responses due to environmental stress often manifest themselves as chemical changes to various photochemical pigments (e.g., chlorophyll, carotenoids, carotene) that control the specific wavelengths at which plant leaves absorb or reflect light [5]. These stress-induced changes are therefore detectable via their effect on pigment response, provided the specific wavelengths at which responses occur can be resolved [6]. Measures of foliar spectra provide information about water content, plant productivity (e.g., chlorophyll), and

overall plant health, and spectral changes can indicate plant stress that may occur from a variety of sources, such as heat, inconsistent soil moisture, and ecological disruptions [7].

While hyperspectral responses have proven effective for the analysis of vegetation conditions, the volume of the data presents a challenge to their use [8]. These challenges have typically been addressed through the use of indices derived from the data [9–11]. Typically, these indices are calculated by combining spectral responses from two or more regions of the spectrum known to be sensitive to various aspects of plant physiology. These band combinations effectively reduce the dimensionality of the data while retaining the spectral information needed for its effective application [10]. For example, chlorophyll responses occur at wavelengths throughout the visible region of the spectrum, so stress-induced effects on chlorophyll can be detected by indices constructed from reflectance in this spectral range [12]. Similarly, responses in the accessory pigments are also detectable in these wavelengths, albeit at more specific wavelengths [5]. Spectral indices sensitive to these pigment responses are widely used in vegetation stress detection [13,14]. Indices further into the near-infrared region have been used for detecting and quantifying water stress in plant leaves [15]. These indices have been applied to the measurement of environmental stress in trees, including drought and increasing temperatures (see, for example, [16,17]). However, these indices have not previously been used as classification features to differentiate between trees based on stress status.

Our research presents a classification methodology for the detection of new or emerging stress in trees using indices derived from hyperspectral data and tests whether existing hyperspectral indices are effective when used as the classification features for this problem. We tested six existing indices—Water Band Index (WBI), Gitelson–Merzlyak B Index (GMB), Normalized Phaeophytization Index (NPQI), Combined Carotenoid/Chlorophyll Ratio Index (CCRI), Photochemical Reflectance Index (PRI), and Red-Edge Chlorophyll Index (Cire)—along with a seventh Test Index—generated as a composite of PRI and Cire—as classification features. Analysis was conducted using data collected from green and white ash trees (*Fraxinus pennsylvanica* Marshall and *Fraxinus americana* L.) with and without emerald ash borer (EAB; *Agrilus planipennis* Fairmaire) infestation to develop a methodology that can be adapted to measure emerging stress from other pathogens or invasive pests in other tree species. Previous related work has focused on the identification of damage or stress symptoms caused by a specific pathogen. In this study, we developed a system of classification that can be applied to the identification of stress symptoms from a range of sources, measurable in trees based on spectral response and, in some cases, detectable prior to the onset of visible symptoms that can be measured through human observation.

By analyzing leaf-level hyperspectral data for the detection of infestation stress in individual trees, based on changes in foliar chemistry, this research expands the applications of indices derived from hyperspectral data in several ways. First, our approach utilizes hyperspectral indices as classification features for categorizing trees based on their condition, infestation status, and health. Second, this technique is focused on detecting new stress in individual trees, separate from existing background stress caused by growing conditions in the trees' environment. This allows for the assessment of the spread or spatial extent of an emerging source of stress, such as a pathogen or invasive pest. In addition, our methodology compiles multiple indices and compares them with each other, both in terms of their efficacy in identifying new plant stress and in terms of their correlation with each other, thereby developing an effective combination to provide a best practice for plant stress detection at the leaf level using hyperspectral indices. Finally, this analysis utilizes data and techniques developed during a multiyear study in order to identify optimal timing during the stages of leaf development through the growing season for the collection of sample data to best identify trees affected by a new source of stress. While the data presented in this analysis pertain to a specific pest affecting two tree species (in this case, EAB in green and white ash trees), these were employed as a test case for developing a methodology that can be adapted for the detection of other stressors in other tree species.

The important goals of this current research are, therefore, to (1) determine the efficacy of the various indices as classification features to classify trees into stress categories based on the detection of spectral changes, (2) identify the optimal combination of hyper-spectral indices sensitive to the early detection of emerging stress in trees, and (3) lay the groundwork for a method of detecting new stress in trees across a range of species by linking foliar chemistry to species properties and monitoring for changes indicative of infestation or other harm. This work addresses two key questions. First, can hyperspectral indices be used effectively as classification features to differentiate trees with differing states of infestation and sources of stress? Second, how and when is this methodology best applied to generate a protocol for the early detection of emerging tree stress?

2. Materials and Methods

To address the research questions outlined above, we identified five groups of trees with varying infestation and stress conditions. Three of the sampling sites were located in Johnson County, Kansas, a suburban area west of Kansas City, Missouri, USA. Emerald ash borer was identified in Johnson County in July 2013; thus, each of these sites was classified as infested. These three infested sites were chosen to represent a variety of growing environments and included a group of trees located along a residential street, one in a recreational sports complex, and one in an urban park. In order to be certain that a comparative sample area would have ash trees that were definitely not EAB-infested, we established two more collection sites in Riley County, Kansas, USA, where monitoring had so far indicated no EAB presence (Bomberger, K, Kansas Department of Forestry, Personal Communication). The first of these two sites consisted of non-infested, mature ash trees located on a university campus that is maintained as an arboretum. These trees were maintained in a more favorable environment and were in a relatively healthy condition at the time of data collection. The second Riley County group also consisted of ash trees located on the campus but in less favorable environments around parking lots and high-traffic areas. These trees were in poorer conditions, including some impacted by endemic fungi, especially *mycosphaerella* spp. This second group was added in order to compare non-infested trees in poorer overall condition with those known to be in a poor condition due to EAB presence. For a more detailed description of the study sites, see [18].

Leaf samples for analysis were collected from each of the five sample site groups described above four times during each of the 2017 through 2019 growing seasons. Collection dates were selected to correspond to four general stages in the foliar season: (1) early leaf stage, (2) mature leaf stage, (3) peak greenness, and (4) late leaf stage. Each of these stages represents variation in the age of the leaf (which affects its internal optical properties, see [6], and tree physiology, e.g., nutrient or water stress). Weather conditions and seasonal differences precluded the exact replication of collection dates between the three years, so efforts were instead made to sample at times consistent with these four stages of leaf development. Sample dates and sizes for each site are summarized in Table 1. Because our method was focused on identifying spectral changes at specific points in leaf development, for any given growing season, collection data were aggregated by leaf development stage across all years of the study rather than by year. To the extent possible, leaf samples were collected from the same trees during each data collection; however, this was not always feasible since some trees died or were removed during the study period. Whenever possible, trees lost from the sample groups were replaced with others from the same area and conforming to similar conditions.

Table 1. Summary of sample collection.

Site Type	Sample Size			Collection Dates ¹			Notes ²
	2017	2018	2019	2017	2018	2019	
Infested, Not Treated (INT)	19	27	16				Residential site, Johnson County
Infested, Treated (IT)	18	17	17	Early: 29 April–6 May Mature: 25–28 June Peak: 26 August–8 September Late: 17–25 October	Early: 9–13 May Mature: 1–9 July Peak: 1–7 September Late: 29–30 September	Early: 13–19 May Mature: 18–25 June Peak: 30 July–6 August Late: 15–18 September	Recreation ctr. site, Johnson County
Recently Infested, Treated (RIT)	9	9	9				Park site, Johnson County
Not Infested, Good Condition (NIG)	15	15	15				Campus site, Riley County
Not Infested, Poorer Condition (NIP)	16	15	15				Campus site, Riley County

¹ Each site was sampled during each of the indicated sampling periods. ² See methods section for additional information about sampling sites.

Sample leaves used for pigment and spectral analyses were collected from random mid-canopy locations within each of the sampled trees using a telescoping pruner, temporarily stored in a cool and dark container, and then transported to the lab for further analysis. Spectral reflectance curves were collected from the adaxial surface of the terminal leaflet of each sample using an ASD-FR spectrometer equipped with a leaf clip with an internal light source. High spectral resolution data are often used to calculate differences or ratio-based indices, which characterize the spectral response in specific wavelengths or spectral regions known to be relevant to leaf structure, physiology, or water content [19]. A wide array of indices have been used for applications of this type, including those related to canopy structure, biochemistry, and physiology [20]. Since one of our research questions addressed the efficacy of using these indices derived from spectral data, we chose seven indices to test. Six of these indices were selected based on a survey of existing literature on the effects of EAB infestation on spectral reflectance [21–23] as well as other types of plant physiological stress more generally [12,15,24,25]. In addition to these existing indices, we added an additional one, called the Test Index (TI). See Table 2 for a detailed description of the indices. The seven indices were calculated from each of the ASD reflectance spectra, using the HSDAR package in R [26].

Table 2. Spectral indices tested.

Index	Definition ¹	Descriptive Purpose of Index ²	Reference
Water Band Index (WBI)	R970/R900	Water stress, general organism stress	[17]
Gitelson–Merzlyak B Index (GMB)	R750/R700	Chlorophyll estimation, leaf senescence	[27]
Normalized Phaeophytization Index (NPQI)	R435–R415/R435+R415	Chlorophyll breakdown and general environmental stress	[28]
Combined Carotenoid/Chlorophyll Ratio Index (CCRI)	((R720–R521)/R521)/((R750–R705)/R705)	Combination of a carotenoid index with a red-edge chlorophyll index	[13]
Photochemical Reflectance Index (PRI)	(R531–R570)/(R531+R570)	General organism stress	[14]
Red-Edge Chlorophyll Index (CIre)	(R750–R705)/R705	Chlorophyll content	[29]
Test Index (TI)	PRI/CIre	Composite of PRI and CIre	This Study

¹ R indicates reflectance in a particular band or wavelength, in nanometers. ² Based on published literature. WBI, GMB, NPQI, CCRI, PRI, and CIre from previous literature on effects of EAB on spectral reflectance. TI new index for this study.

Feature selection was used to determine which of the seven hyperspectral indices (Table 2) were best suited for the classification step and which might be redundant. To determine which spectral indices to use, principle component analysis (PCA) was run separately on all five iterations of the data (four individual collections and the combined collection). Indices that were correlated in PCA space were considered redundant and were candidates for elimination from further analysis. The degree of association between the various indices was determined by examining the loading matrices for each of the collections. Indices with high negative or positive loadings on a particular principal component were considered correlated and, in those cases, the index that was most closely correlated with a particular component was retained for further analysis. This feature selection method was chosen over other methods (such as simple correlation) because it allowed the strength of association between the various indices to be determined in multivariate space [30].

Classification was performed using the gradient-boosted decision tree [31], with categorical feature splits made using the information gain criteria based on entropy [32]. Decision tree classification was chosen for this application because it is a non-parametric technique relatively insensitive to noisy input data [33] and is relatively robust in the face of small sample sizes [34]. This last criterion was important in this application, since some of the datasets classified were small, especially when single collections were classified.

3. Results

Graphical presentations of the first two components of the PCA are shown in Figure 1 for the entire dataset (all four collection time periods) and in Figure 2 for each of the four individual collection periods. In these figures, vectors for highly correlated variables lie along the same line, with positively correlated variables pointing in the same direction and negatively correlated variables pointing in opposite directions. Vectors for variables that were uncorrelated would be orthogonal to each other. For example, the CCRI and TI variables for all collections (see Figure 1) would be highly negatively correlated, and both would be nearly uncorrelated with the Water Band Index (WBI).

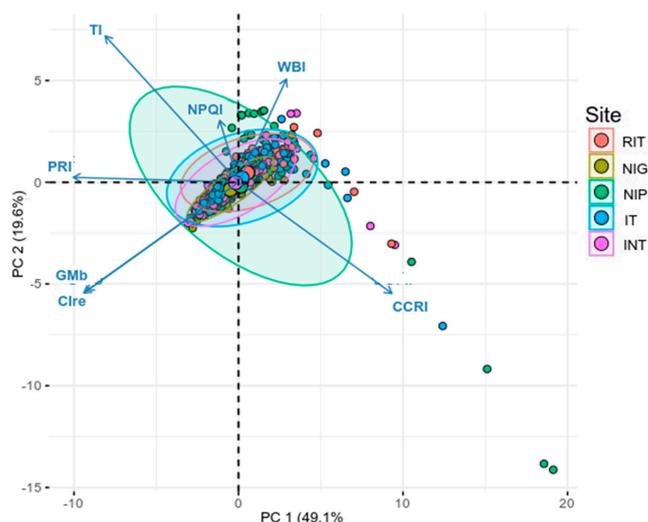


Figure 1. Bivariate plot of first two principal components of the spectral index data for all four collection periods combined. The vectors associated with each of the seven input spectral variables are shown by the blue arrows in the diagram. Sites are defined in Table 1. Percentage of variance for each component is indicated in parentheses.

Examination of the PCA plots in Figures 1 and 2 shows that while some indices (e.g., CIre and GMB) were consistently correlated throughout all four collection time periods, most of the indices showed some variation in their correlation patterns, which made the determination of redundant variables difficult. To evaluate redundancy more thoroughly in

the data, we used the loading matrices output by the PCA analysis, summarized in Table 3. To simplify the process, we used only the first four of the seven components generated by the analysis. These four components accounted for more than 90% of the total variance in the dataset and therefore adequately summarized the information within the dataset. The PRI, CIRE, CCRI, and GMB indices all showed similar loading on the first principle component throughout all iterations of the dataset. The absolute loading values of these four indices typically varied from about 0.44 to 0.48, which indicated a modest level of significant correlation; however, the consistency of these associations across the entire range of the data indicated redundancy in these four indices.

Table 3. Loading matrices for principle components 1 through 4 for each collection period and for all collections combined.

Early Leaf					Mature Leaf				
Index	PC1	PC2	PC3	PC4	Index	PC1	PC2	PC3	PC4
PRI	0.46	0.38	−0.08	0.09	PRI	0.44	0.41	0.05	0.29
CIRE	0.49	−0.30	0.15	−0.15	CIRE	0.53	−0.26	0.07	−0.11
CCRI	−0.49	0.11	0.01	−0.16	CCRI	−0.41	0.33	0.06	−0.40
WBI	0.11	0.07	−0.81	−0.55	WBI	0.04	−0.06	−0.97	0.12
NPQI	−0.01	0.45	0.52	−0.69	NPQI	0.25	0.44	−0.19	−0.72
GMB	0.48	−0.29	0.17	−0.16	GMB	0.52	−0.20	0.09	−0.20
TI	0.24	0.68	−0.08	0.36	TI	0.13	0.65	0.04	0.41
Peak Greenness					Late Leaf				
Index	PC1	PC2	PC3	PC4	Index	PC1	PC2	PC3	PC4
PRI	−0.41	−0.41	−0.07	0.18	PRI	−0.48	−0.04	−0.09	−0.02
CIRE	−0.47	0.19	0.26	−0.30	CIRE	−0.44	0.33	−0.10	0.38
CCRI	0.43	−0.11	−0.37	−0.28	CCRI	0.44	0.29	−0.29	0.21
WBI	0.28	−0.21	0.84	0.25	WBI	0.06	−0.68	−0.05	0.73
NPQI	0.11	−0.65	0.12	−0.66	NPQI	−0.12	−0.25	−0.90	−0.28
GMB	−0.47	0.17	0.15	−0.38	GMB	−0.45	0.34	−0.11	0.32
TI	−0.33	−0.53	−0.21	0.40	TI	−0.41	−0.41	0.26	−0.32
All Collections									
Index	PC1	PC2	PC3	PC4					
PRI	−0.48	0.02	0.21	0.01					
CIRE	−0.44	−0.40	−0.12	−0.30					
CCRI	0.45	−0.41	0.17	−0.14					
WBI	0.14	0.38	−0.37	−0.82					
NPQI	−0.05	0.23	0.87	−0.35					
GMB	−0.45	−0.41	−0.06	−0.25					
TI	−0.39	0.54	−0.12	0.21					

Among the remaining three indices (WBI, NPQI, and TI), WBI and NPQI consistently loaded strongly on specific components. WBI loaded very significantly on the third component in the early leaf stage, mature leaf stage, and peak greenness stage, the fourth component in the late leaf stage, and the composite of all collection stages. On the other hand, NPQI loaded most strongly on the second component (often a negative loading) for all collection periods except the late leaf stage and the composite of all collections, where it loaded strongly on component 3. Not only were WBI and NPQI not well correlated with each other, based on their loadings, but each was typically the only index that loaded strongly onto a particular component. Bivariate plots of the components (Figures 1 and 2) further showed that these two indices were not well correlated with each other or with the

four correlated indices described above. Similarly, the TI tended to load onto a specific component (in this case, component 2) and, like WBI and NPQI, it tended to be the only index loaded onto that component, though not as strongly or consistently as the other indices.

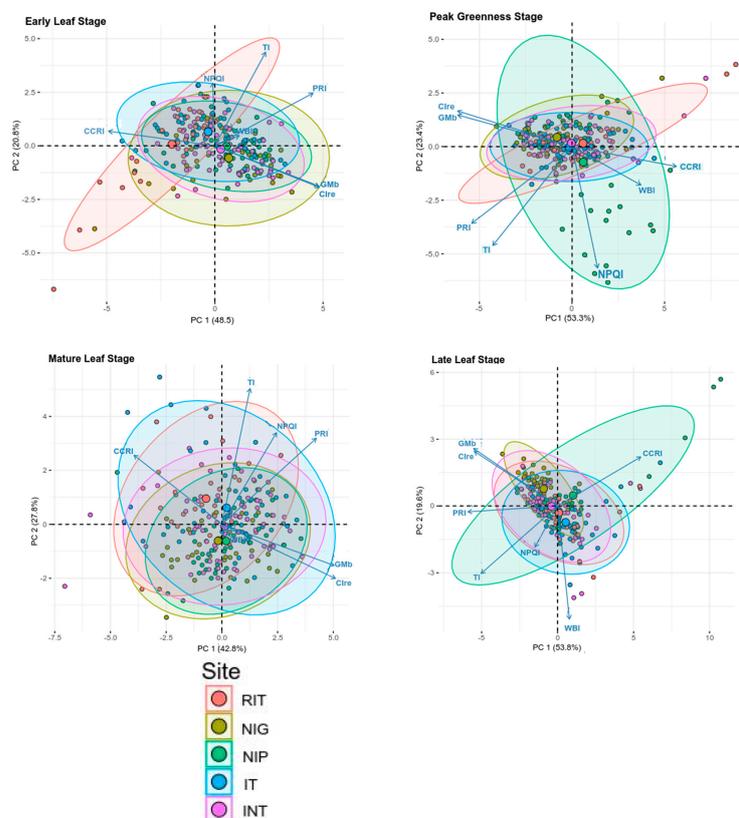


Figure 2. Bivariate plots of first two principal components of the spectral index data for each collection period (see Table 1 for collection dates). The vectors associated with each of the seven input spectral variables are shown by the blue arrows in the diagram. Sites are defined in Table 1. Percentage of variance for each component is indicated in parentheses.

Because principal component analysis provides a more detailed assessment of the relationships between indices (compared to simple correlation), it was useful for identifying the most informative combination of features [35,36]. Based on the PCA results, we chose to retain four indices for use in the subsequent classification step: WBI, NPQI, CCRi, and TI. We used these four least correlated hyperspectral indices, rather than the components, because we wanted our classification method to be based on directly measured features, rather than on the synthetic features resulting from the eigenanalysis of the original indices. Further analysis and classification of the five sites was therefore performed using only these four hyperspectral indices.

Five separate classifications were performed using the four selected features: one combining all of the collections (see Table 4) and one for each of the four data collection periods (see Table 5a–d). Numbers in Table 4 through 5d (error matrices) represent the number of instances classified by category, with the accuracy and Kappa value displayed at the bottom of each table. We separated the analyses in this way to test whether there were seasonal variations in the effectiveness of our classification method and, if so, determine at what times in the active growing season the classification methods worked best. Prior to each of the classifications, the data were split into two equal-sized parts: one for training and the other for validation. The classification results and error matrices presented in this section were derived only from the validation set. All classifications were done using the gradient-boosted decision tree (GBDT) method implemented in the GBM library of the R analysis language [37]. Error matrices and accuracy statistics were calculated using functions from the R Caret library [38].

Table 4. Error matrices from all collection periods combined.

		Reference				
		NIG	NIP	IT	INT	RIT
Prediction	NIG	37	3	5	10	1
	NIP	4	32	7	8	4
	IT	6	8	43	8	2
	INT	12	5	11	42	2
	RIT	4	1	3	9	22
	Accuracy	60.9%				
Kappa	0.505					

Table 5. Error Matrices for each collection. a. Error matrices for collection 1 (early leaf stage). b. Error matrices for collection 2 (mature leaf stage). c. Error matrices for collection 3 (peak greenness stage). d. Error matrices for collection 4 (late leaf stage). See Table 1 for dates of each collection.

a. Error Matrices, Collection 1 (Early Leaf Stage)						
		Reference				
		NIG	NIP	IT	INT	RIT
Prediction	NIG	12	1	0	0	0
	NIP	2	9	2	2	0
	IT	2	1	12	1	1
	INT	2	3	0	11	1
	RIT	0	2	0	3	4
	Accuracy	67.6%				
Kappa	0.598					
b. Error Matrices, Collection 2 (Mature Leaf Stage)						
		Reference				
		NIG	NIP	IT	INT	RIT
Prediction	NIG	11	0	1	2	0
	NIP	1	12	1	1	0
	IT	1	2	12	1	3
	INT	0	0	3	11	1
	RIT	1	0	0	1	7
	Accuracy	73.6%				
Kappa	0.668					
c. Error Matrices, Collection 3 (Peak Greenness Stage)						
		Reference				
		NIG	NIP	IT	INT	RIT
Prediction	NIG	8	0	2	3	0
	NIP	1	7	0	5	2
	IT	1	2	9	4	1
	INT	1	1	1	16	0
	RIT	0	1	2	2	4
	Accuracy	60.3%				
Kappa	0.490					

Table 5. Cont.

d. Error Matrices, Collection 4 (Late Leaf Stage)						
		Reference				
		NIG	NIP	IT	INT	RIT
Prediction	NIG	16	1	1	3	0
	NIP	0	10	4	3	0
	IT	1	1	21	3	1
	INT	2	2	3	24	1
	RIT	1	2	6	2	4
	Accuracy	67%				
Kappa		0.573				

Using the combined results from all four collection periods, the overall classification accuracy was 60.9%, with a Kappa value of 0.505. The Kappa metric quantifies the improvement of a classification over random assignment [39]. A value like the one reported here suggests that, in this case, the results were significantly improved (vs. random assignment), but that the overall effectiveness of the classification was modest at best.

The classification results from each of the four leaf development stages (not combined) showed substantial variation in accuracy among these stages of the growing season. The best results tended to occur either earlier in the growing season or near its end. Collections 1 and 2 (early leaf stage and mature leaf stage) returned accuracies of about 68% and 74%, respectively (Table 5a,5b), both of which were better than the results obtained using data from the combined collections. Collection 4 (late leaf stage) also returned about 67% accuracy (Table 5d), although its Kappa value indicated somewhat less improvement over random assignment compared to either the early leaf stage (Table 5a) or all collections combined (Table 4). The only collection period for which the accuracy did not exceed that of the combined classification was collection 3 (peak greenness stage), whose overall accuracy value was only slightly greater than 60% and whose Kappa value indicated less than 50% improvement over random assignment (Table 5c).

To test whether there were significant differences between the various classification results, we compared the results for each of the four leaf development time periods using the method of Foody [40]. In this method, pair-wise z-scores were calculated:

$$z = \frac{\frac{X_i}{n_i} - \frac{X_j}{n_j}}{\sqrt{p(1-p)\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \tag{1}$$

where

$$p = \frac{X_i + X_j}{n_i + n_j} \tag{2}$$

In this analysis, X represented the number of correctly classified cases in each sample, while n was the sample size (number of observations) for each collection. The results of these comparisons showed that the classification results for each of the sampling periods were significantly separable from at least one of the other leaf development time periods, and, with the exception of the mature leaf period, all were separable from the results of the combined collection (Table 6). The period of peak greenness was significantly separated from all of the other collection periods, whereas the early and mature periods were not separable. Since the greatest classification accuracy occurred at the early and mature leaf stages, this suggests that data collected at these two times have the greatest potential for successfully discriminating infested from non-infested trees. Furthermore, the comparatively poorer performance of the classifier at the mature leaf stage compared to all

other stages indicates that this is the time when the classification is least likely to correctly identify infested versus non-infested trees. Classification during the late-leaf period would also be more effective than at the mature stage, although less so than at times earlier in the growing season.

Table 6. Z-scores and *p*-values for pair-wise comparisons of classification accuracy between each collection period and the combined collection. Collection periods are defined in Table 1.

	Combined	Early Leaf Stage	Mature Leaf Stage	Peak Greenness	Late Leaf Stage
Combined	-	-	-	-	-
Early Leaf Stage	1.07 (0.142)	-	-	-	-
Mature Leaf Stage	2.05 * (0.020)	1.18 (0.119)	-	-	-
Peak Greenness	0.084 (0.467)	1.45 ** (0.074)	2.79 * (0.003)	-	-
Late Leaf Stage	1.17 (0.121)	0.21 (0.492)	1.31 (0.095)	1.83 * (0.034)	-

* = significant at $p \leq 0.05$. ** = significant at $p \leq 0.10$.

4. Discussion

In this research, we addressed two questions: can hyperspectral indices be used to classify trees with differing states of infestation and sources of stress and when (in the leaf development cycle) can this methodology best be applied to generate a protocol for the early detection of emerging tree stress? Our research was able to provide answers to both of these questions, while also demonstrating their importance to the problem and their relationship to each other.

The spread of emerald ash borer in North America provided a useful test case for developing and analyzing our classification method. Native to Asia, the invasive emerald ash borer (*Agrilus planipennis* Fairmaire) is believed to have arrived in North America during the mid-1990s, and was first identified in southeastern Michigan in 2002 [41]. The adult insect, approximately 8.5 to 12.5 mm in length and bright green in color, bores into the tree and lays eggs in the cambium [42]. After hatching, the larvae eat their way through the phloem and xylem, eventually tunneling around the tree, under the bark, creating a girdling effect that slowly cuts off the flow of moisture and nutrients to higher portions of the tree [43]. As upper branches are deprived of water and nutrients, they become defoliated and, over time, large sections of the crown may be devoid of leaves during the growing season. These easily visible symptoms of disease do not usually become apparent until six years after initial infestation and, because they result from significant, irreversible damage to the tree's internal transport system, it is generally too late to save the tree by the time they appear [44].

The progression of the pathology of EAB infestation illustrates the importance of early detection and the value of developing a classification tool to differentiate between affected and non-affected trees and track the spread of infestation. However, this disease progression also makes emerald ash borer useful as a representative pathogen because the disruption to the flow of nutrients and moisture through the tree is similar to the types of damage and stress caused by other tree diseases and invasive pests [45]. Dutch elm disease, for example, results from a fungal infection, generally introduced by a boring insect, that spreads from tree to tree and leads to rapid wilting and death due to the blocked flow of moisture and nutrients [46]. The development of more sensitive tools for the early detection of infestation and the identification of affected individuals is therefore important for a variety of trees and tree diseases [47], and similarities in the types of stress produced in impacted trees suggest that our method could be beneficial when adapted to other species and pathogens.

Using EAB as a model pathogen system, our results show that hyperspectral indices can be used to sort trees into infestation categories. We evaluated seven such indices

and found that a subset of four of them was able to group trees into infestation classes using a boosted-gradient decision tree classifier. The four specific indices that proved most effective also shed some light on which leaf physiological properties are affected by disease infestation and how these effects determine the infestation class of an individual tree. WBI responds to water stress and was the only one of the four indices that was not linked in some way to leaf pigment response. It was also the only one that was derived mostly from leaf reflectance in the SWIR spectral region. Its inclusion indicates that leaf water status is affected by infestation and the resulting stress on the organism. NPQI, on the other hand, is associated with chlorophyll function and general environmental stress. The inclusion of NPQI in the final predictive model is consistent with previous analyses of the spectral effects of woody pathogen infestation [48]. It is noteworthy that of the non-composite indices we tested that are closely linked to chlorophyll (GMb, CIre, and NPQI, see Table 2), NPQI was the most useful and also the one related to chlorophyll degradation, rather than just to chlorophyll content [28]. This suggests that it is the condition of the leaf chlorophyll, rather than its concentration in the leaf, that is affected by pathogen presence and therefore determines its infestation category. Furthermore, it is notable that this index is computed from reflectance on the shorter end of the visible spectrum, emphasizing the importance of spectral response across the spectrum in assessing the effects of pathogen stress.

The two remaining indices we used in the final classification model, CCRI and TI, were both composites that combined other indices. This suggests that the impacts of pathogen stress manifest themselves spectrally in multiple ways and that indices that take this into account are more effective than those that target only one effect. In the case of CCRI, this index reflects multiple pigment effects by including the impacts of stress on both carotenoids and chlorophylls [25]. It also includes the influence of the red edge [49], a spectral region closely associated with vegetation conditions [50]. TI, which we developed for this study, was a composite of red edge index, sensitive to chlorophyll content, and PRI, a very effective indicator of general organismal stress [51]. Again, the fact that these indices worked better when incorporated into a composite indicates that the combined effects of pathogen stress provide more effective diagnostic features for classification, creating a more sensitive technique for the early detection of emerging stress.

Among the indices that were not included in the final protocol (PRI, GMb, and CIre), their elimination was based in part on the decreased effectiveness of single factor measurements, the need to eliminate redundancies from correlations with other indices, and the lower rates of reliability as an indicator of infestation stress. In some cases, these issues were addressed by the use of composite indices, and they also demonstrated the importance of identifying the optimal timing for sample collection within the leaf cycle stages. For example, previous research indicates that PRI demonstrates some efficacy in differentiating between infested and non-infested trees [18]. However, the effectiveness of PRI as an indicator varied across the foliar season and was not consistent from year to year. The ability of PRI to detect manifestations of stress was incorporated into the composite TI, and reliability and consistency were improved by developing an optimal timing model for data collection.

Seasonality played an important role in influencing the effectiveness of hyperspectral indices for detecting emerging stress. By tying our data collection calendar to the stages of leaf development through the growing season, we were able to mitigate the impact of the variability of environmental conditions from one year to another. This also helped to identify the optimal timing for measuring new or emerging sources of stress, such as pathogen or pest infestation, and separate them from ongoing background stress in the environment, such as drought or heat stress. Our analysis indicates that data collection was most effective at early and late points in the foliar season, when emerging infestation stress had measurable impacts on reflectance indices, distinct from measurements in healthy trees and therefore most useful in classifying trees based on infestation status. Data collected during the hottest part of summer, when all groups of trees were most stressed from heat and moisture, proved least effective for classifying tree infestation status. When trees were most stressed by environmental conditions, they showed the least differentiation in their

response to other stresses. Because the timing of peak heat and moisture conditions varies from year to year, the stages of leaf development—early leaf stage, mature leaf stage, peak greenness, and late leaf stage—present a more reliable guide for timing the collection of diagnostic data. Data from samples collected at the early and mature leaf stages were most effective for successfully identifying infested versus non-infested trees. Data from the mature stage proved least effective for differentiating infested from non-infested trees. The identification of this timing protocol provides an important component in the development of a more sensitive technique for the early detection of emerging stress in trees.

5. Conclusions

The major contribution of this study to the existing body of research is the use of hyperspectral indices as classification features. Previous work using hyperspectral indices for tree research, as reflected in the literature, has utilized these data as indicators of tree health or stress but has not employed them for the classification of trees as infested or non-infested. Our development of this method of analysis significantly expands the usefulness of hyperspectral data in both tree pathology research and forest management in that it demonstrates the potential to identify which trees have been affected by a new pathogen or emerging source of stress within a geographic area that has been impacted by an invasive pest or other new stressor. While sources of stress previously studied—such as drought, heat, or soil conditions—are likely to impact most or all trees within a region, invasive pathogens impact trees individually and spread spatially, making the early detection and identification of impacted trees, as well as the tracking of disease spread, crucial to addressing a new stressor and potentially limiting damage.

Our research demonstrates the effectiveness of hyperspectral indices as classification features to classify trees into stress categories based on the detection of spectral changes and identifies an effective combination of hyperspectral indices sensitive to the early detection of emerging stress in trees, along with a timing protocol for data collection. Previous research, including some of our own, has focused on the use of indices derived from hyperspectral data for measuring levels of stress or injury in trees based on spectral indicators of health. This current work instead presents a method for using hyperspectral indices as classification features to differentiate between affected and non-affected trees in areas impacted by a new or spreading pathogen. It further lays out a timing protocol for separating emerging infestation stress from ongoing background stress.

Our intention in this analysis was to demonstrate a potential approach to detecting and classifying disease-infested trees, using ash species and EAB as model systems. It may be that applying this methodology to other species of pest or tree would require the selection of new indices, although testing this current model on another infestation system would be instructive. This work lays the foundation of a strategy for detecting and differentiating new stress in trees, across a range of species and a variety of pathogens, by linking foliar chemistry to species properties, monitoring for changes indicative of infestation or other harm, and utilizing indices as classification features for categorizing the status of affected trees.

Author Contributions: Conceptualization, L.M.M.; methodology, L.M.M. and D.G.G.; validation, L.M.M., D.G.G. and W.P.W.III; formal analysis, L.M.M., D.G.G. and W.P.W.III; investigation, L.M.M., D.G.G. and W.P.W.III; resources, L.M.M., D.G.G. and W.P.W.III; data curation, D.G.G.; writing—original draft preparation, L.M.M. and D.G.G.; writing—review and editing, L.M.M., D.G.G. and W.P.W.III; visualization, L.M.M., D.G.G. and W.P.W.III; supervision, L.M.M.; project administration, L.M.M.; funding acquisition, D.G.G. and W.P.W.III. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a University Small Research Grant from Kansas State University and the Kansas Forest Service.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors acknowledge the assistance of Kim Bomberger, Kansas Forest Service, and Craig Shafer, Johnson County Parks and Recreation Department.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Lassalle, G.; Fabre, S.; Credoza, A.; Hédacq, R.; Bertoni, G.; Dubucq, D.; Elger, A. Application of PROSPECT for Estimating Total Petroleum Hydrocarbons in Contaminated Soils from Leaf Optical Properties. *J. Hazard. Mater.* **2019**, *377*, 409–417. [[CrossRef](#)] [[PubMed](#)]
- Pontius, J.; Martin, M.; Plourde, L.; Hallett, R. Ash Decline Assessment in Emerald Ash Borer-Infested Regions: A Test of Tree-Level, Hyperspectral Technologies. *Remote Sens. Environ.* **2008**, *112*, 2665–2676. [[CrossRef](#)]
- Pu, R. *Hyperspectral Remote Sensing: Fundamentals and Practices*, 1st ed.; CRC Press: Boca Raton, FL, USA; Taylor & Francis: Abingdon, UK, 2017; ISBN 978-1-315-12060-7.
- Shukla, A.; Kot, R. An Overview of Hyperspectral Remote Sensing and Its Applications in Various Disciplines. *IRA-JAS* **2016**, *5*, 85. [[CrossRef](#)]
- Blackburn, G.A. Hyperspectral Remote Sensing of Plant Pigments. *J. Exp. Bot.* **2006**, *58*, 855–867. [[CrossRef](#)] [[PubMed](#)]
- Huang, J.; Wei, C.; Zhang, Y.; Blackburn, G.A.; Wang, X.; Wei, C.; Wang, J. Meta-Analysis of the Detection of Plant Pigment Concentrations Using Hyperspectral Remotely Sensed Data. *PLoS ONE* **2015**, *10*, e0137029. [[CrossRef](#)]
- Curran, P.J. Remote Sensing of Foliar Chemistry. *Remote Sens. Environ.* **1989**, *30*, 271–278. [[CrossRef](#)]
- Chutia, D.; Bhattacharyya, D.K.; Sarma, K.K.; Kalita, R.; Sudhakar, S. Hyperspectral Remote Sensing Classifications: A Perspective Survey. *Trans. GIS* **2016**, *20*, 463–490. [[CrossRef](#)]
- Song, G.; Wang, Q. Developing Hyperspectral Indices for Assessing Seasonal Variations in the Ratio of Chlorophyll to Carotenoid in Deciduous Forests. *Remote Sens.* **2022**, *14*, 1324. [[CrossRef](#)]
- Jin, J.; Wang, Q. Informative Bands Used by Efficient Hyperspectral Indices to Predict Leaf Biochemical Contents Are Determined by Their Relative Absorptions. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 616–626. [[CrossRef](#)]
- Gitelson, A.; Vina, A.; Arkebauer, T.; Rundquist, D.; Keydan, G.; Leavitt, B.; Verma, S. Remote Estimation of Leaf Area Index and Green Leaf Biomass in Maize Canopies. *Geophys. Res. Lett.* **2003**, *30*, 1248–1251. [[CrossRef](#)]
- Blackburn, G.A. Spectral Indices for Estimating Photosynthetic Pigment Concentrations: A Test Using Senescent Tree Leaves. *Int. J. Rem. Sens.* **1998**, *19*, 657–675. [[CrossRef](#)]
- Zhou, X.; Huang, W.; Kong, W.; Ye, H.; Dong, Y.; Casa, R. Assessment of Leaf Carotenoids Content with a New Carotenoid Index: Development and Validation on Experimental and Model Data. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *57*, 24–35. [[CrossRef](#)]
- Gamon, J.; Serrano, L.; Surfus, J.S. The Photochemical Reflectance Index: An Optical Indicator of Photosynthetic Radiation Use Efficiency across Species, Functional Types, and Nutrient Levels. *Oecologia* **1997**, *112*, 492–501. [[CrossRef](#)]
- Penuelas, J.; Baret, F.; Filella, I. Semi-Empirical Indices to Assess Carotenoids/Chlorophyll a Ratio from Leaf Spectral Reflectance. *Photosynthetica* **1995**, *31*, 221–230.
- Datt, B. Remote Sensing of Water Content in Eucalyptus Leaves. *Aust. J. Bot.* **1999**, *47*, 909. [[CrossRef](#)]
- Penuelas, J.; Filella, I.; Biel, C.; Serrano, L.; Save, R. The Reflectance at the 950–970 Nm Region as an Indicator of Plant Water Status. *Int. J. Rem. Sens.* **1993**, *14*, 1887–1905. [[CrossRef](#)]
- Moley, L.M.; Goodin, D.G.; Winslow, W.P. Leaf-Level Spectroscopy for Analysis of Invasive Pest Impact on Trees in a Stressed Environment: An Example Using Emerald Ash Borer (*Agilus Planipennis* Fairmaire) in Ash Trees (*Fraxinus* spp.), Kansas, USA. *Environments* **2022**, *9*, 42. [[CrossRef](#)]
- Ustin, S.L.; Gitelson, A.A.; Jacquemoud, S.; Schaepman, M.; Asner, G.P.; Gamon, J.A.; Zarco-Tejada, P.J. Retrieval of Foliar Information about Plant Pigment Systems from High Resolution Spectroscopy. *Remote Sens. Environ.* **2009**, *113*, S67–S77. [[CrossRef](#)]
- Roberts, D.; Roth, K.; Perroy, R. Hyperspectral Vegetation Indices. In *Hyperspectral Remote Sensing of Vegetation*; CRC Press: Boca Raton, FL, USA, 2011; pp. 309–328. ISBN 978-1-4398-4537-0.
- Hu, B.; Li, J.; Wang, J.; Hall, B. The Early Detection of the Emerald Ash Borer (EAB) Using Advanced Geospatial Technologies. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 213. [[CrossRef](#)]
- Zhang, K.; Hu, B.; Robinson, J. Early Detection of Emerald Ash Borer Infestation Using Multisourced Data: A Case Study in the Town of Oakville, Ontario, Canada. *J. Appl. Remote Sens.* **2014**, *8*, 083602. [[CrossRef](#)]
- Pontius, J.; Hanavan, R.P.; Hallett, R.A.; Cook, B.D.; Corp, L.A. High Spatial Resolution Spectral Unmixing for Mapping Ash Species across a Complex Urban Environment. *Rem. Sens. Environ.* **2017**, *199*, 360–369. [[CrossRef](#)]
- Carter, G.A.; Knapp, A.K. Leaf Optical Properties in Higher Plants: Linking Spectral Characteristics to Stress and Chlorophyll Concentration. *Am. J. Bot.* **2001**, *88*, 677–684. [[CrossRef](#)] [[PubMed](#)]
- Zhou, X.; Huang, W.; Zhang, J.; Kong, W.; Casa, R.; Huang, Y. A Novel Combined Spectral Index for Estimating the Ratio of Carotenoid to Chlorophyll Content to Monitor Crop Physiological and Phenological Status. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *76*, 128–142. [[CrossRef](#)]

26. Lehnert, L.W.; Meyer, H.; Obermeier, W.A.; Silva, B.; Regeling, B.; Bendix, J. Hyperspectral Data Analysis in R: The Hsdar Package. *J. Stat. Softw.* **2019**, *89*. [[CrossRef](#)]
27. Gitelson, A.; Merzlyak, M.N. Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus Hippocastanum* L. and *Acer Platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *J. Plant Physiol.* **1994**, *143*, 286–292. [[CrossRef](#)]
28. Barnes, J.D.; Balaguer, L.; Manrique, E.; Elvira, S.; Davison, A.W. A Reappraisal of the Use of DMSO for the Extraction and Determination of Chlorophylls a and b in Lichens and Higher Plants. *Env. Exp. Bot.* **1992**, *32*, 85–100. [[CrossRef](#)]
29. Gitelson, A.A.; Viña, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote Estimation of Canopy Chlorophyll Content in Crops. *Geophys. Res. Lett.* **2005**, *32*, L08403. [[CrossRef](#)]
30. Song, F.; Guo, Z.; Mei, D. Feature Selection Using Principal Component Analysis. In Proceedings of the 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, Yichang, China, 12 November 2010; Volume 1, pp. 27–30.
31. Shendryk, Y.; Rossiter-Rachor, N.A.; Setterfield, S.A.; Levick, S.R. Levick Leveraging High-Resolution Satellite Imagery and Gradient Boosting for Invasive Weed Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4443–4450. [[CrossRef](#)]
32. De'ath, G.; Fabricius, K.E. Classification and Decision Trees: A Powerful yet Simple Technique for Ecological Data Analysis. *Ecology* **2000**, *81*, 3178–3192. [[CrossRef](#)]
33. Funkenberg, T.; Binh, T.T.; Moder, F.; Dech, S. The Ha Tien Plain—Wetland Monitoring Using Remote-Sensing Techniques. *Int. J. Remote Sens.* **2014**, *35*, 2893–2909. [[CrossRef](#)]
34. Kavzoglu, T.; Colkesen, I. A Kernel Functions Analysis for Support Vector Machines for Land Cover Classification. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 352–359. [[CrossRef](#)]
35. Tahir, N.M.; Hussain, A.; Samad, S.A.; Ishak, K.A.; Halim, R.A. Feature Selection for Classification Using Decision Tree. In Proceedings of the 2006 4th Student Conference on Research and Development, Shah Alam, Malaysia, 27 June 2006; pp. 99–102.
36. Sun, L.; Xu, J.; Yin, Y. Principal Component-Based Feature Selection for Tumor Classification. *Bio-Med. Mater. Eng.* **2015**, *26*, S2011–S2017. [[CrossRef](#)]
37. Greenwell, B.; Boehmke, B.; Cunningham, J. Gbm: Generalized Boosted Regression Models 2020. R Package Version **2020**, 2.
38. Kuhn, M. Caret: Classification and Regression Training 2020. R Package Version **2021**, 3.
39. Landis, J.; Koch, G. An Application of Hierarchical Kappa-Type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics* **1977**, *33*, 363–374. [[CrossRef](#)]
40. Foody, G.M. Thematic Map Comparison. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 627–633. [[CrossRef](#)]
41. Haack, R.A.; Jendak, E.; Houping, L.; Marchant, K.R.; Petrice, T.R.; Poland, T.M.; Ye, H. The Emerald Ash Borer: A New Exotic Pest in North America. *Newsl. Mich. Entomol. Soc.* **2002**, *47*, 1–5.
42. Volkovitsh, M.G.; Orlova-Bienkowskaja, M.J.; Kovalev, A.V.; Bieńkowski, A.O. An Illustrated Guide to Distinguish Emerald Ash Borer (*Agrilus planipennis*) from Its Congeners in Europe. *For. Int. J. For. Res.* **2019**, *93*, 316–325. [[CrossRef](#)]
43. Herms, D.A.; McCullough, D.G. Emerald Ash Borer Invasion of North America: History, Biology, Ecology, Impacts, and Management. *Annu. Rev. Entomol.* **2014**, *59*, 13–30. [[CrossRef](#)] [[PubMed](#)]
44. Poland, T.M.; McCullough, D.G. Emerald Ash Borer: Invasion of the Urban Forest and the Threat to North America's Ash Resource. *J. For.* **2006**, *104*, 118–124. [[CrossRef](#)]
45. Schlarbaum, S.E.; Hebard, F.; Spaine, P.C.; Kamalay, J.C. Three American Tragedies: Chestnut Blight, Butternut Canker, and Dutch Elm Disease. In *Proceedings of the Exotic Pests of Eastern Forests Conference Proceedings*; U.S. Forest Service and Tennessee Exotic Pest Plant Council: Nashville, TN, USA, 1998; pp. 45–54.
46. Webber, J.F. Experimental Studies on Factors Influencing the Transmission of Dutch Elm Disease. *For. Syst.* **2004**, *13*, 197–205.
47. Potter, C.; Harwood, T.; Knight, J.; Tomlinson, I. Learning from History, Predicting the Future: The UK Dutch Elm Disease Outbreak in Relation to Contemporary Tree Disease Threats. *Phil. Trans. R. Soc. B* **2011**, *366*, 1966–1974. [[CrossRef](#)]
48. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [[CrossRef](#)] [[PubMed](#)]
49. Horler, D.N.H.; Dockray, M.; Barber, J. The Red Edge of Plant Leaf Reflectance. *Int. J. Remote Sens.* **1983**, *4*, 273–288. [[CrossRef](#)]
50. Reflectance at the Red Edge as a Sensitive Indicator of the Damage of Trees and Its Correlation to the State of the Photosynthetic System. Nasr, H.N. (Ed.) In *Proceedings of the Image Understanding for Aerospace Applications*; SPIE: Munich, Germany, 1991; p. 131.
51. Zarco-Tejada, P.J.; González-Dugo, V.; Williams, L.E.; Suárez, L.; Berni, J.A.J.; Goldammer, D.; Fereres, E. A PRI-Based Water Stress Index Combining Structural and Chlorophyll Effects: Assessment Using Diurnal Narrow-Band Airborne Imagery and the CWSI Thermal Index. *Remote Sens. Environ.* **2013**, *138*, 38–50. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.