

Context-dependent performance of Savitzky-Golay smoothing in hyperspectral Forest species classification: A Bayesian hierarchical assessment

Dávid Abriha^{a,*}, Gergő Diószegi^a, Lilla Kovács^a, Péter Burai^{b,c}, Szilárd Szabó^a

^a Department of Physical Geography and Geoinformatics, University of Debrecen, Egyetem tér 1, Debrecen, Hungary

^b EnviroSense Hungary Ltd., 4032 Debrecen, Péchy u. 46, Hungary

^c Remote Sensing Centre, University of Debrecen, Böszörményi Út 138, 4032 Debrecen, Hungary

ARTICLE INFO

Keywords:

Hyperspectral imagery
Savitzky-Golay filtering
Tree species classification
Bayesian hierarchical modelling

ABSTRACT

Savitzky-Golay (SG) filtering can mitigate the ‘salt-and-pepper’ effect in pixel-wise classifications of hyperspectral imagery, but its optimal parametrization across diverse forest types remains unclear. We evaluated 81 SG parameter combinations (Window: 3–19, Order: 1–17) on species-level classifications across three structurally distinct Hungarian forests (Gemenc, Mecsek, Ságvár) using a Bayesian hierarchical model to quantify the effects of forest site, tree species, their interaction, and SG parameters on overall accuracy. Our analysis revealed three principal findings. First, forest site was the primary driver of SG performance, while individual species showed no significant effect; only one forest-species interaction (*Mecsek::pedunculate oak*) proved significant. Second, 72 of 81 parameter combinations were statistically equivalent, indicating substantial parameter flexibility. Third, the combination Window 5 and Order 1 (b₅:1) emerged as the most robust performer, while excessive variance reduction systematically degraded accuracy, particularly in the most diverse forest (Mecsek). We conclude that SG filter performance is governed by forest-level structural and community characteristics rather than species identity, and the wide equivalence of most parameters offers practical flexibility, with b₅:1 recommended as a robust default. Beyond these specific findings, this study demonstrates the value of adopting a hierarchical modelling framework –here, the Bayesian Hierarchical Model –for evaluating pre-processing techniques, offering an adaptable analytical template for future research across diverse remote sensing applications.

1. Introduction

Accurate identification of tree species is critical for forest management, underpinning applications in biodiversity conservation (Jia et al., 2024; Shang and Chisholm, 2014), hazard and risk management (Cho et al., 2010; Fassnacht et al., 2014), and invasive species monitoring (Burai et al., 2011; Piironen et al., 2018; Zhao et al., 2022). Beyond these immediate management needs, tracking the flourishing, decline, or migration of tree species serves as a critical indicator of climate change impacts (Dyderski et al., 2018; Van Houtven et al., 2019; Woodall et al., 2010).

Advances in remote sensing have greatly improved the accuracy of mapping tree species. Airborne hyperspectral imaging, which offers detailed spatial and spectral resolution, is widely used to differentiate tree species as an alternative to traditional, time-consuming field

surveys (Likó et al., 2024; Modzelewska et al., 2020; Shi et al., 2018). In forested areas with many spectrally similar species, the high spectral resolution of hyperspectral remote sensing offers a clear advantage over multispectral imagery. This finer resolution increases the likelihood of resolving the unique spectral profile of each species, thereby improving classification accuracy (Ghiyam et al., 2013; Richter et al., 2016; Shi et al., 2018).

However, the high resolution of airborne hyperspectral imagery also introduces challenges, including multicollinearity across hundreds of spectral bands, increased noise, and computational complexity (Camps-Valls and Bruzzone, 2005; Melgani and Bruzzone, 2004; Richter et al., 2016; Zhang et al., 2022). These issues are well-recognized and are routinely addressed by techniques such as dimensionality reduction (e.g., Principal Component Analysis - PCA, Independent Component Analysis - ICA) and noise reduction (e.g., Minimum Noise Fraction

* Corresponding author.

E-mail address: abriha.david@science.unideb.hu (D. Abriha).

<https://doi.org/10.1016/j.ecoinf.2026.103914>

Received 2 March 2026; Received in revised form 30 June 2026; Accepted 1 July 2026

Available online 3 July 2026

1574-9541/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

-(MNF) (Hycza et al., 2018; Likó et al., 2022; Liu and Wu, 2018; Shebl et al., 2023). In contrast, recent deep learning approaches, such as multi-stream convolutional neural networks and spectral feature fusion frameworks, aim to learn robust spectral-spatial representations directly from high-dimensional hyperspectral data by exploiting complex inter-band and contextual dependencies (Li et al., 2020; Li et al., 2022).

Classification of high spatial resolution images also presents an often-overlooked problem: the salt-and-pepper (SAP) effect. This phenomenon, common in pixel-wise classification, results in scattered misclassifications that degrade both map accuracy and visual interpretability (Abriha et al., 2023; Hirayama et al., 2019; Liu and Wu, 2018). A common approach to mitigating the SAP effect is post-classification filtering (e.g., Gaussian, median, majority filters), which smooths isolated misclassified pixels (Vargas-Muñoz et al., 2017; Yulianto et al., 2021; Zhu, 2013). An alternative strategy is pre-smoothing, where spatial filters (e.g., Gaussian, moving average, median) are applied to the image *prior* to classification. The goal of pre-smoothing is to increase pixel homogeneity and reduce spatial heterogeneity, leading to a more robust classification (Li et al., 2023; Liang et al., 2021; Lukin et al., 2013; ZhiYong et al., 2018).

Another powerful pre-smoothing technique is the Savitzky-Golay (SG) filtering, which performs convolution by fitting a polynomial of a specified order to a moving window (Savitzky and Golay, 1964). While moving average or median filters may be more effective for noise reduction alone, SG's key advantage is its ability to preserve original spectral characteristics—a critical property for applications requiring a balance between noise suppression and spectral detail retention (Kawala-Sterniuk et al., 2020; Ruffin et al., 2008). However, the performance of SG filtering is highly dependent on its two key parameters—Window size and polynomial Order—and its effectiveness across structurally and compositionally diverse forest types remains poorly understood.

Recent studies have become increasingly focused on using complex algorithms to address the high dimensionality and structural complexity of forest remote sensing data. For instance, researchers have employed advanced deep learning frameworks, such as Vision Transformers (Nepi et al., 2025) and multi-temporal convolutional neural networks (Lin et al., 2024; Veras and Ferreira, 2022) to identify tree species. Similarly, advanced statistical optimisation approaches, such as Bayesian optimisation, are increasingly being used to tune machine learning models and identify optimal hyperspectral feature subsets (Song and Wang, 2023). While these methods often deliver strong predictive performance, they also shift the analytical focus towards increasingly complex and computationally intensive classification frameworks. By contrast, considerably less attention has been devoted to understanding how fundamental pre-processing techniques affect the statistical structure of hyperspectral data before classification. In particular, robust and interpretable frameworks for evaluating how SG filtering interacts with ecological context remain limited. Most existing studies treat pre-processing as a fixed step rather than a tunable component whose optimization depends on site-specific ecological characteristics. This gap is critical because the effectiveness of spatial smoothing is likely mediated by forest structure, species composition, and phenological state-factors that vary considerably across study sites. The present study therefore shifts the focus from classifier complexity to ecologically informed input data optimisation. Specifically, we apply SG filtering as a pre-processing step prior to Random Forest classification, and use the resulting overall accuracy as the response variable in a Bayesian Hierarchical Random Model (BHRM). This allows us to quantify how ecological context (forest site, species, and their interaction) and methodological choices (Window size and SG Order) jointly influence

classification performance.

This study had three primary objectives: (i) to quantify the influence of ecological factors (forest type, tree species, and their statistical interaction) and methodological factors (81 SG parameter combinations of Window size and Order) on SG filter performance across three heterogeneous forests; (ii) to identify the most robust SG parameter combinations, defined by their consistent performance across diverse forest conditions; and (iii) to demonstrate the value of a Bayesian hierarchical random effects model as an analytical framework for evaluating pre-processing techniques in remote sensing, capable of handling the complex, nested structure of ecological data.

2. Methods

This study integrates airborne hyperspectral remote sensing data with spectral preprocessing, machine learning classification, and Bayesian hierarchical modelling to evaluate the effects of Savitzky-Golay (SG) filtering on tree species classification performance. The overall methodological workflow consists of (i) data acquisition from three forest study sites, (ii) preprocessing and SG-based spectral smoothing using multiple parameter combinations, (iii) Random Forest classification of both original and filtered hyperspectral datasets, and (iv) statistical evaluation of classification performance using a Bayesian hierarchical framework to quantify the effects of ecological context and preprocessing choices (Fig. 1.).

2.1. Study sites

The analyses were conducted in three different forest areas in the southern Transdanubian region of Hungary: Ságvár, Mecsek, and Gemenc, varying in size at 800, 210, and 70 ha, respectively (Fig. 2). The surveyed areas displayed high ecological diversity encompassing a variety of tree species. While native species such as European beech (*Fagus sylvatica*), pedunculate oak (*Quercus robur*), and Norway spruce (*Picea abies*) continue to dominate the areas, invasive trees such as black locust (*Robinia pseudoacacia*) or common hackberry (*Celtis occidentalis*) were observed in all locations. Naturally occurring hybrids were also present in Gemenc and Ságvár.

2.2. Datasets

The imagery was acquired using two airborne hyperspectral sensors: the AISA FENIX 1 K for the Mecsek and Gemenc study areas, and the AISA Kestrel 10 for the Ságvár area. All datasets were processed to a uniform spatial resolution of 1 m. Acquisition dates differed among the study areas, and differences in sensor spectral configurations resulted in varying numbers of spectral bands and spectral ranges used in the analyses (Table 1). Non-forested areas were masked, and only tree species were included in the analysis.

For the classification, reference data were gathered through field surveys and examination of hyperspectral imagery. Despite the relatively small size of the study sites, they show a high degree of heterogeneity. Specifically, we collected reference data from 8 species in the Ságvár forest, 7 in Mecsek, and 5 in Gemenc (Tables 2–4).

2.3. Savitzky-Golay filtering

The SG filter is a versatile data smoothing tool that balances noise reduction with the preservation of important features in the data. By fitting polynomials within sliding windows, it can effectively remove noise while preserving the integrity of the underlying signal (Chen, 2004; Jinhu et al., 2010). We applied the SG method implemented in

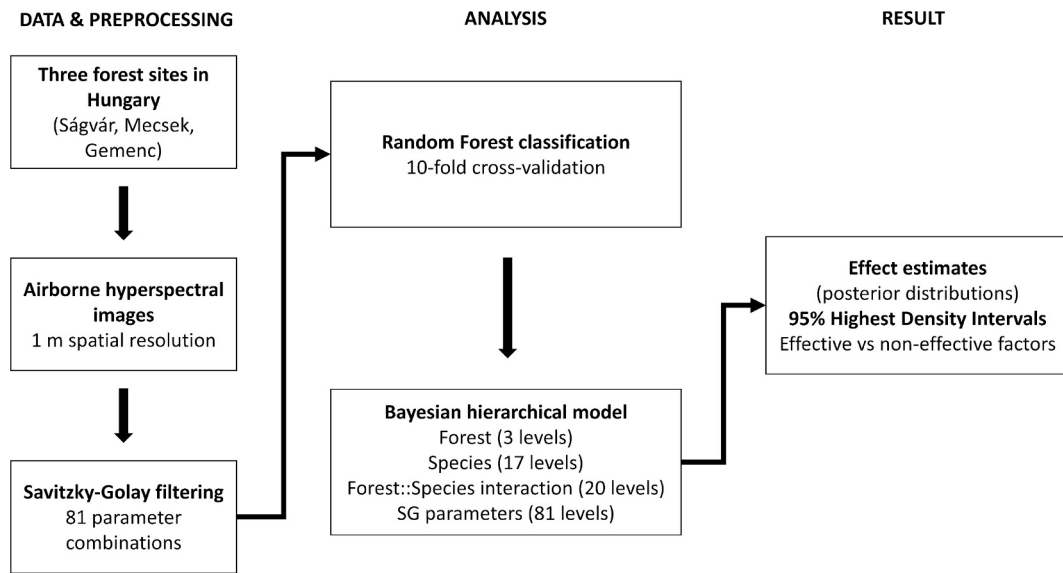


Fig. 1. Overview of the methodological workflow.

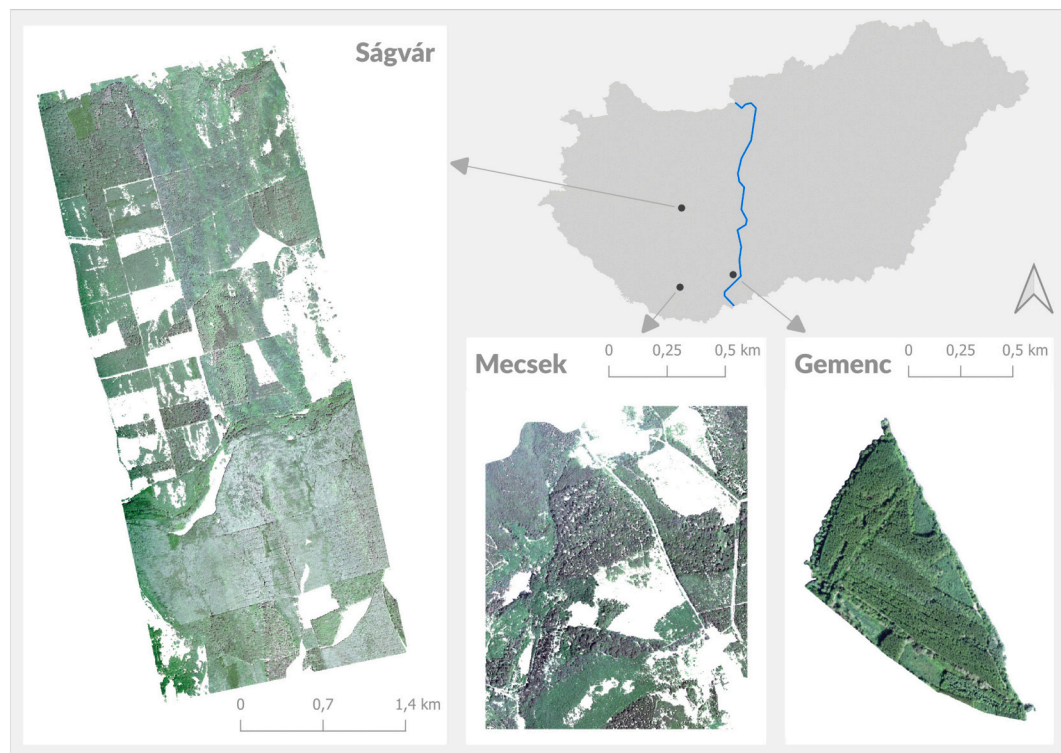


Fig. 2. Location and forested areas of the study sites.

Table 1

The acquisition dates and spectral resolutions of the hyperspectral images.

Dataset	Acquisition date	Spectral resolution
Ságvár	2022-08-12	431–1010 nm (170 bands)
Mecsek	2016-06-29	442–1080 nm (158 bands)
Gemenc	2018-09-12	472–974 nm (147 bands)

Python using the *pysptools* package. There are two important criteria when choosing the parameters: (i) the window size must be specified as an odd integer, and (ii) the order of the polynomial used in the filtering

process must be less than the window size minus one. To investigate the relationship between the parameters, we analyzed permutations ranging from 3 to 19 for moving window sizes and 1 to 17 for polynomial orders. This approach allowed a thorough examination of the impact of varying these parameters on the effectiveness and performance of the filtering process, resulting in an additional 81 filtered images available for classification.

2.4. Image classification

Although there are several classification methods, we applied the Random Forest (RF), an ensemble learning methodology that has been

Table 2

Scientific name, common name and abbreviation of the tree species collected in the Ságvár region, with an asterisk (*) denoting invasive species.

Scientific name	Common name	Abbreviation
<i>Robinia pseudoacacia</i> *	black locust	BL
<i>Celtis occidentalis</i> *	common hackberry	CH
<i>Pinus sylvestris</i>	Scotch pine	SP
<i>Tilia tomentosa</i>	silver linden	SL
<i>Juglans nigra</i>	black walnut	BW
<i>Alnus glutinosa</i>	common alder	CA
<i>Populus x euramericana</i>	Canadian poplar	CP
<i>Quercus cerris</i>	turkey oak	TO

Table 3

Scientific name, common name and abbreviation of the tree species collected in the Mecsek region, with an asterisk (*) denoting invasive species.

Scientific name	Common name	Abbreviation
<i>Robinia pseudoacacia</i> *	black locust	BL
<i>Ailanthus altissima</i> *	tree of heaven	TH
<i>Tilia tomentosa</i>	silver linden	SL
<i>Acer pseudoplatanus</i>	sycamore maple	SM
<i>Quercus robur</i>	pedunculate oak	PO
<i>Quercus petraea</i>	sessile oak	SO
<i>Picea abies</i>	Norway spruce	NS

Table 4

Scientific name, common name and abbreviation of the tree species collected in the Gemenc region, with an asterisk (*) denoting invasive species.

Scientific name	Common name	Abbreviation
<i>Acer negundo</i> *	box elder	BE
<i>Pinus nigra</i>	black pine	BP
<i>Quercus robur</i>	pedunculate oak	PO
<i>Salix alba</i> cv <i>Populus x euramericana</i> cv. Agathe-F	OP-229 poplar	OP
<i>Populus x euramericana</i> cv. Pannónia	H-490/3 poplar	HP

widely used in the field of tree species classification and is well known for its robustness (Cutler et al., 2007; Dalponte et al., 2013; Immitzer et al., 2012; Qin et al., 2022). We applied RF in both the original and filtered images, employing the *Scikit-learn* package of Python (Pedregosa et al., 2018). The objective of this research is to investigate the differences between the original images and those filtered using the SG method with varying parameters, rather than to emphasize the importance of hyperspectral optimization. Consequently, our classification experiments stand for methodological considerations, and models were trained using the default parameters specified in the package's documentation without the aim of getting the most accurate outcomes. Thus, we avoided the differences among the models due to different hyperparameters.

2.5. Bayesian hierarchical random model

To isolate the effects of forest site, tree species, and SG parameters on classification accuracy, we employed a Bayesian hierarchical model. This framework is well-suited to our data structure, where observations are nested within species and forests, and where we needed to account for both main effects and interactions simultaneously (Gelman et al., 2013).

The response variable was overall accuracy (OA) derived from Random Forest classification of both SG-filtered and non-smoothed images, calculated from each fold of a 10-fold cross-validation. The categorical predictors were: *sav* (SG parameter combination with 81 levels, representing the preprocessing choice), *Forest* (the three forest sites), *Species* (17 tree species), and the *Forest::Species* interaction (20

levels). This specification partitions variance in OA into components attributable to: (i) ecological context (*Forest*, *Species*, and their interaction), (ii) methodological choices (SG parameter combinations), and (iii) residual error. By treating OA as the response, we effectively model how pre-processing choices interact with ecological context to influence classification performance. Because our predictors are categorical and hierarchically nested (species within forests), we treated them as random effects (Faraway, 2016; Gomez-Rubio, 2020). We have disabled partial pooling for the *Forest*, *Species*, *sav*, and *Forest::Species* interaction terms, meaning that no information was shared across categories within each variable. This approach treats each category independently, avoiding potential shrinkage that could mask true differences between groups (Kurz, 2021). The model was specified as:

$$OA \sim Forest + Species + Forest :: Species + sav + \epsilon$$

where *Forest::Species* denotes the interaction term and ϵ represents residual error. Each random effect was assigned an independent Gaussian distribution with a mean of zero. Following standard recommendations for hierarchical models, we assigned weakly informative priors to all variance parameters –conservative enough to prevent extreme estimates but flexible enough to let the data dominate the inference (Gelman et al., 2013; Gomez-Rubio, 2020).

The model was fitted using Integrated Nested Laplace Approximation (INLA) via the R-INLA package (Rue et al., 2009). INLA provides a computationally efficient alternative to Markov Chain Monte Carlo (MCMC) sampling, which is essential for complex hierarchical structures like ours (R Core Team, 2022). A posterior predictive check confirmed that the model adequately captured patterns in the observed data (Kruschke, 2014; Kurz, 2021).

Following Bayesian conventions, we considered an effect to be *effective* (i.e., statistically meaningful) if the 95% Highest Density Interval (HDI) of its posterior distribution did not contain zero. This terminology—*effective* versus *non-effective*—is used throughout the study.

3. Results

3.1. Classification accuracies across study sites and Savitzky-Golay variates

The initial analysis of the classification performance across 81 SG parameter combinations indicates a clear stratification in baseline and smoothed accuracies driven by the specific study sites (Fig. 3). A notable trend across the entire study is that the vast majority of SG variates outperform the non-smoothed baseline. The Gemenc forest exhibited the most substantial response to preprocessing. Accuracy improvements often reached 5–8%, achieving a peak accuracy of about 95.8%, which represents a 10.4% absolute increase over the baseline.

To further examine the interaction between site-specific characteristics and methodological settings, deviations from the original baseline were mapped across all window sizes and polynomial orders (Fig. 4).

This analysis reveals that for all forest sites, performance remains stable or improves when using smaller window sizes regardless of the polynomial order (Fig. 4). However, a significant decline in performance—reaching losses as high as 31% in Gemenc and Mecsek—is evident when large window sizes (e.g., 13 to 19) are paired with low polynomial orders. This combination led to oversmoothing and the subsequent loss of diagnostic spectral features necessary for species differentiation.

The textural effects associated with different SG variates remain consistent across all study sites. The true-colour composites show part of the Gemenc forest as a representative example (Fig. 5).

The original imagery exhibits high spectral variability at the crown scale (Fig. 5a). The optimized b_5::1 configuration reduces this fine-scale variability while preserving distinct crown boundaries and canopy structure (Fig. 5b). In contrast, the lowest-performing b_19::1 variate reduces crown-scale spatial heterogeneity, resulting in increasingly

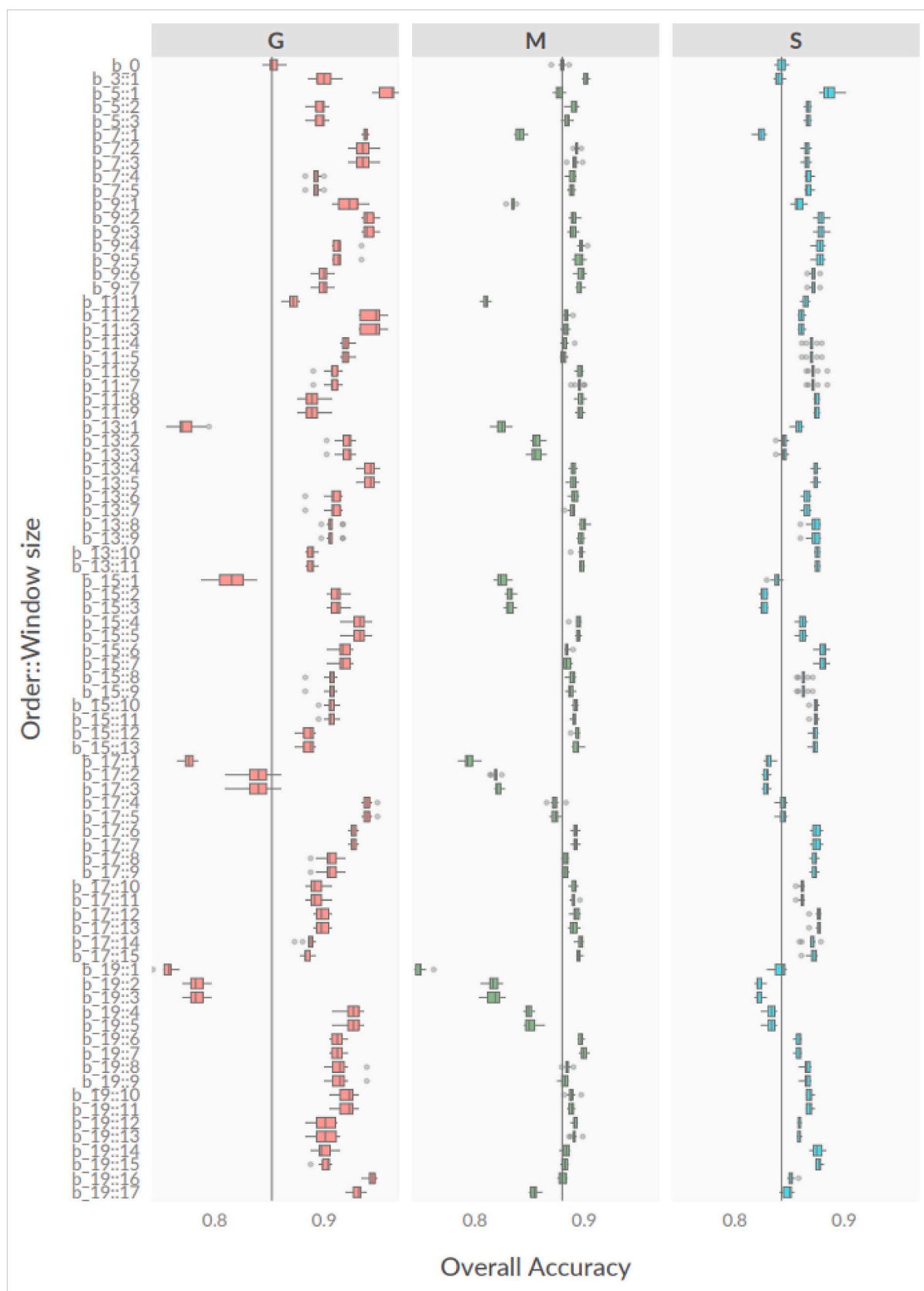


Fig. 3. Comparison of overall classification accuracies across 81 Savitzky-Golay filter variates for the G = Gemenc, M = Mecsek, and S=Ságvár forest study areas. The vertical grey line indicates the baseline median accuracy achieved using the non-smoothed image (b_0), while the y-axis denotes the variates identified by the interaction (denoted by the:: symbol) between the moving window size and polynomial order.

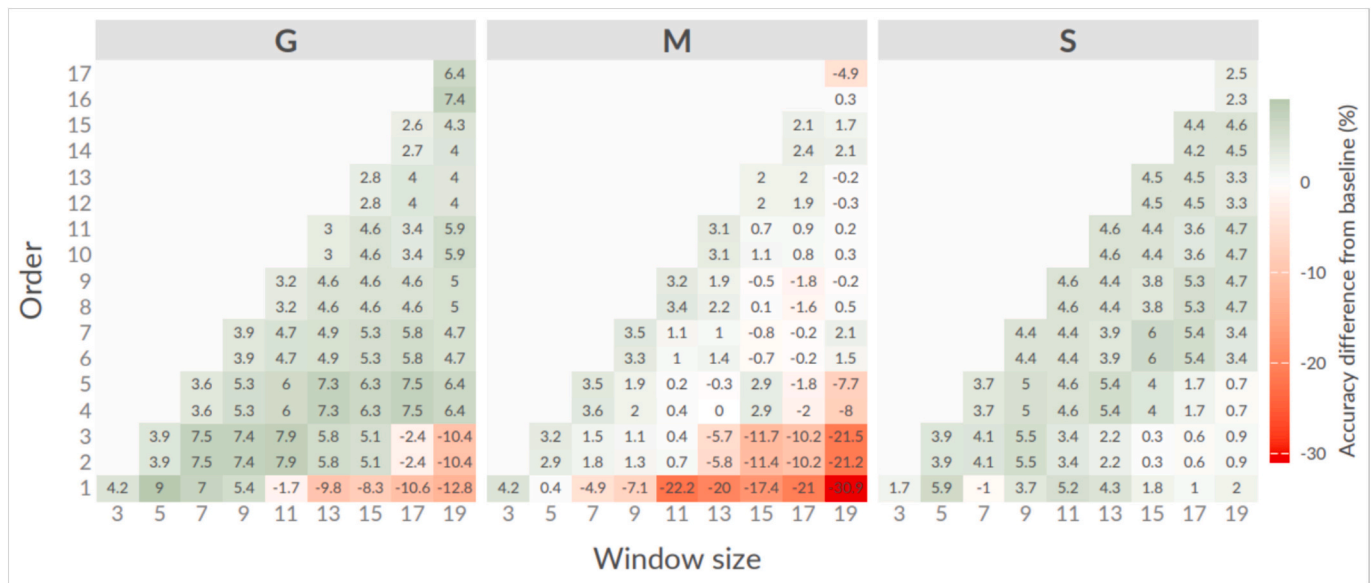


Fig. 4. Performance deviations from the baseline accuracy across G = Gemenc, M = Mecsek, and S=Ságvár forests.

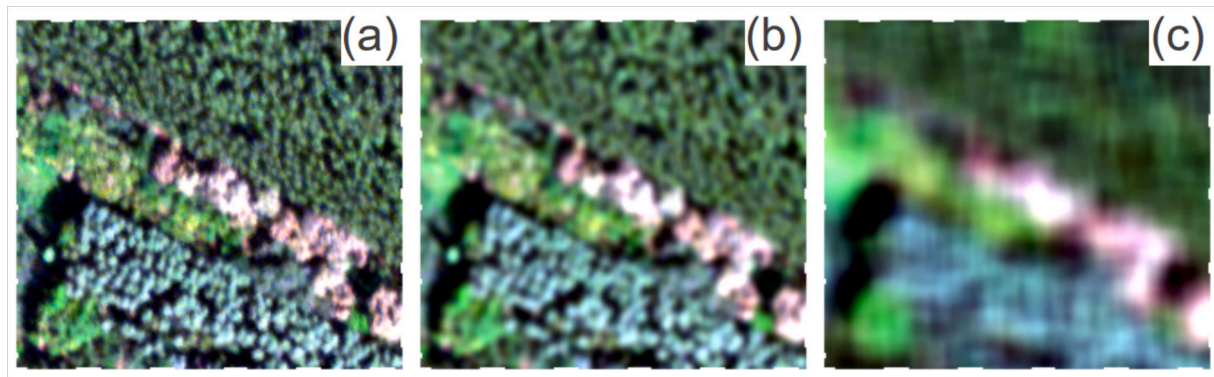


Fig. 5. Visual impact of Savitzky-Golay (SG) filtering on canopy texture in a subset of the Gemenc study area. (a) Original (non-smoothed) imagery; (b) highest-performing SG variate (b_5::1); (c) lowest-performing SG variate (b_19::1).

uniform reflectance patterns across adjacent canopy features. Therefore, the observed decline in classification performance for this parameter configuration is consistent with the reduced preservation of crown-level detail evident in the imagery.

To further quantify the relationship between filter intensity and classification effectiveness, we analyzed the interaction between accuracy and the Variance Reduction Ratio (VRR), calculated as the ratio of the variance in the SG smoothed spectrum to the variance of the original signal (Fig. 6).

The three sites exhibit different responses when subjected to increasing variance reduction. Ságvár shows the narrowest VRR range and the most stable behavior, while Gemenc tolerates substantially higher VRR before accuracy degradation becomes evident. In contrast, Mecsek is the most sensitive to smoothing; its classification performance declines rapidly once VRR exceeds moderate levels. Across all sites, classification accuracy peaks within a low-to-intermediate VRR range, suggesting that modest variance reduction improves model performance, while higher VRR values lead to systematic accuracy losses. As

the VRR increases beyond the optimal range, the accuracy declines across all sites. This is accompanied by a substantial increase in the inconsistency of smoothing among the classes. This pattern is most pronounced in Mecsek forest, where high VRR configurations coincide with large performance losses and elevated interspecific disparity. The VRR analysis independently identified the highest-performing configurations for each site: Window size 5 and Order 1 (b_5::1) for both Gemenc and Ságvár, and Window size 3 and Order 1 (b_3::1) for Mecsek (Fig. 6).

While most configurations outperform the baseline (Fig. 3), the high degree of variability in gains across different geographic contexts—influenced by ecological traits such as tree species (Fig. 7)—necessitates a more robust analytical framework. These distributions consequently serve as the empirical foundation for the BHRM, which is used in the following sections to statistically isolate the most robust parameter influences.

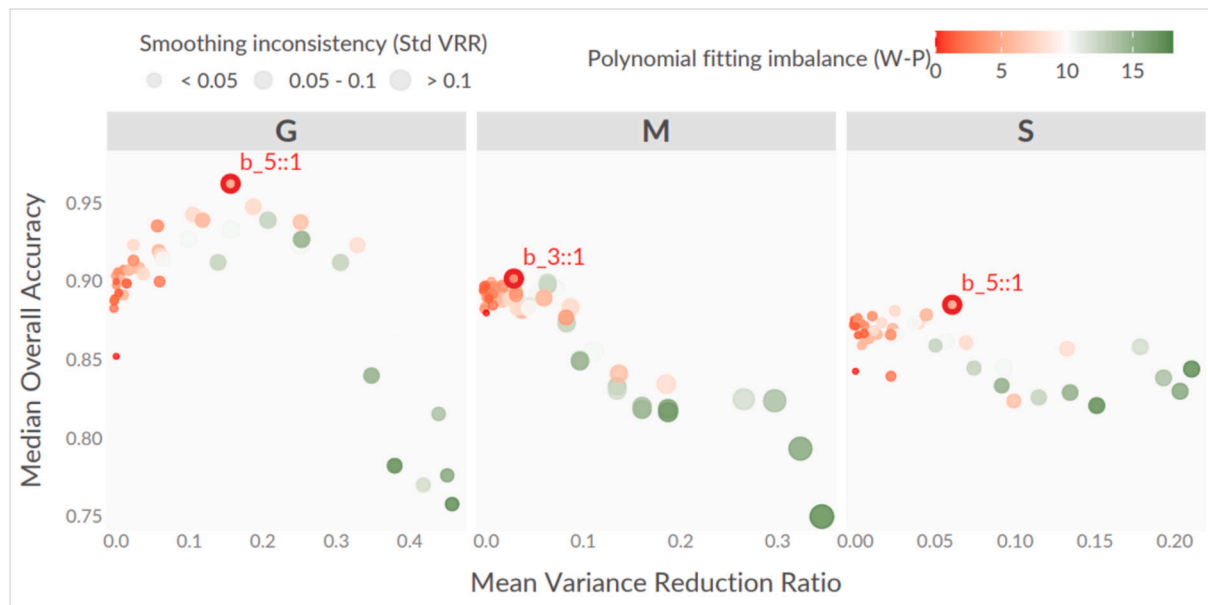


Fig. 6. The impact of Savitzky-Golay filtering intensity on classification performance across G = Gemenc, M = Mecsek, and S=Ságvár forests. Marker colour represents the polynomial fitting imbalance (window size - polynomial order), while marker size reflects the interspecific smoothing disparity across classes (standard deviation). The highest performing configuration for each site is annotated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Effects of forest and species variables

The model results indicate that *Forest* has an effective impact on smoothing accuracy: zero effect lied outside the HDIs for all three forests (Fig. 8) suggesting that the structural and compositional differences between the Gemenc, Mecsek, and Ságvár forests significantly influence smoother performance. In contrast, when we examined the data independently of *Forest*, the individual *Species* effects were all found to be non-effective, as their HDIs all contained zero (Fig. 8). This result underscores the importance of having modelled the *Forest::Species* interaction, as the influence of species appears to be context-dependent on the forest environment.

3.3. Effects of the forest-species interaction

Analysis of the *Forest::Species* interaction revealed that, out of 20 combinations, only a single interaction was effective: specifically, that of Mecsek Forest with pedunculate oak (Fig. 9). The posterior distributions for all other interactions contained zero within their 95% HDIs, indicating non-effective, context-specific effects.

3.4. Effects of the smoother parameter interaction (window size-order)

A primary objective was to quantify the impact of the SG smoother parameter combinations on OA and based on the posterior distributions, we distinguished effective from non-effective parameter combinations. Nine of the 81 combinations that were evaluated proved effective, but the vast majority (72) were non-effective, meaning that their impact on accuracy was not statistically different from zero. (Fig. 10). Critically, all 9 effective parameter combinations had a negative effect, with their 95% HDIs entirely below zero, indicating they consistently resulted in lower OAs.

3.5. Identification of the most robust smoother

The 72 non-effective smoothers were statistically equivalent in our Bayesian framework, as their HDIs all included zero (Fig. 11). Among this group, the combination of Window size 5 and Order 1 (denoted

b_5::1) was identified as the most robust based on its superior point-estimate performance, exhibiting the highest posterior mean overall accuracy (Fig. 11). It is important to emphasize that while b_5::1 is optimal by this metric, its non-effectivity means it was statistically indistinguishable from the other 71 smoothers in the group, all of which provided a similarly robust, negligible influence on accuracy.

4. Discussion

This study employed a Bayesian hierarchical model to assess the impact of Savitzky-Golay (SG) smoothing on classification accuracy across three distinct forest sites. Specifically, we examined how site-specific factors (e.g., species composition, phenological stage encapsulated in the *Forest* attribute), tree species (*Species*), and their interaction (*Forest::Species*) modulated the effect of varying SG filter parameters (Window size and Order) on overall classification accuracy.

The analysis revealed that the effectiveness of SG smoothing is highly context-dependent, with its success hinging on an interaction between filter parameters and site-specific ecological characteristics. Across the three forest sites, most SG-smoothed datasets yielded a measurable improvement in overall accuracy compared to the non-smoothed baseline, although the magnitude of this effect varied considerably (Fig. 3). The most pronounced response was observed for Gemenc, where classification accuracy increased by up to 10.4%. This substantial gain is likely attributable to the forest's phenological state at the time of acquisition (September 12). The onset of leaf senescence is known to induce substantial changes and increase variability in leaf-level reflectance properties, with previous work demonstrating systematic alterations in spectral reflectance due to biochemical and structural degradation in deciduous tree species (Boyer et al., 1988). In this context, SG filtering likely mitigates fine-scale, intra-crown spectral variability, yielding a more stable crown-level spectral representation that aligns with the assumptions of the Random Forest classifier and enhances species discrimination.

In contrast, the Mecsek site exhibited significant variability among species, with performance often declining under more aggressive filtering. In these diverse canopies, large-window smoothing appears to over-smooth spectral signatures, resulting in the loss of subtle diagnostic



Fig. 7. Distribution of species-level accuracy from 10-fold cross-validation. Each panel corresponds to a specific Savitzky-Golay filter parameter combination (Window size::Order). The first panel (marked with an asterisk, *) represents the observed, non-smoothed data; the remaining 81 panels represent distinct parameter combinations. Within a panel, points are grouped by forest (G = Gemenc, M = Mecsek, S=Ságvár) along the y-axis and scattered according to accuracy on the x-axis, with each point representing an individual tree species. The number of trees (points) per forest (in each panel) is: Gemenc (50), Mecsek (70), Ságvár (80).

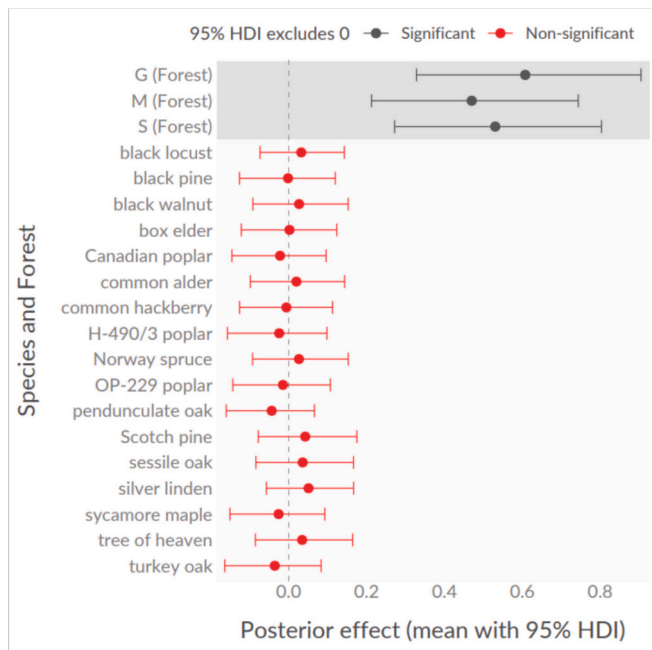


Fig. 8. Parameter estimates from the Bayesian Hierarchical Regression Model (BHRM). Error bars represent the 95% highest density intervals (HDI) of the posterior distributions for the three forest stands (top, highlighted with darker grey background) and the 17 tree species. Dots indicate the posterior mean. The vertical grey dashed line at zero represents the mean of the cross-validated non-smoothed data (i.e., no effect). Forest stands are abbreviated as G (Gemenc), M (Mecsek), and S (Ságvár). Species names are given in full.

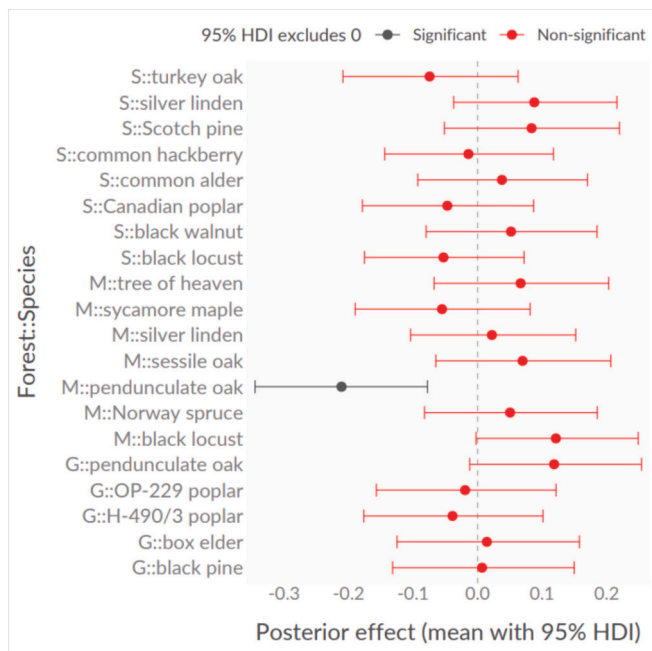


Fig. 9. Posterior estimates of the forest-species interaction. Error bars represent the 95% highest density intervals (HDI) of the posterior distributions, with posterior means shown as dots. The vertical grey dashed line marks the null effect (zero). Forest and species are separated by a double colon (e.g., G::black pine). Only the M::pendunculate oak interaction shows a significant negative effect, as its 95% HDI does not cross zero.

features necessary for accurate differentiation. This pattern matches the observed visual degradation of crown-level detail (Fig. 5c): excessive smoothing produces overly uniform reflectance patterns, effectively obscuring the biologically meaningful canopy heterogeneity that constitutes the ‘signal’ for species discrimination. This performance decline aligns with the findings of Vaiphasa (2006), who used Jeffries-Matusita distance to demonstrate that spectral smoothing can inadvertently reduce the statistical separability between vegetation classes. Collectively, these trends demonstrate that classification performance is governed not solely by the magnitude of variance reduction, but by the balance between noise suppression and the preservation of class-specific spectral variability.

4.1. Convergent evidence from VRR and BHRM analyses

The VRR analysis independently identified the same optimal SG configurations as the Bayesian hierarchical model. For Gemenc and Ságvár, the combination Window size 5 and Order 1 (b₅::1) emerged as the highest-performing variate, while for Mecsek, Window size 3 and Order 1 (b₃::1) was optimal (Fig. 6). This convergence between a heuristic measure of filtering intensity and a formal Bayesian inference framework provides strong corroborating evidence for our findings. The consistency across two independent analytical approaches—one based on variance reduction ratios and the other on posterior distributions from a hierarchical model—reinforces the robustness of b₅::1 as a reliable default configuration. The slight deviation observed for Mecsek (b₃::1) reflects the greater sensitivity of this diverse forest to variance reduction, consistent with our broader conclusion that forest-level context governs smoothing performance.

4.2. Ecological drivers of smoother performance

Results demonstrated that *Forest* was a primary driver of SG smoother accuracy, as evidenced by the significant effects for all three study sites (Fig. 8). In contrast, *Species* alone showed no significant effect (Fig. 8), suggesting that species-specific traits were not the dominant factor. This implies that broader structural, and community-level characteristics encapsulated by the *Forest* variable—rather than species composition alone—were the principal source of performance variation.

While no previous study has conducted as comprehensive an evaluation of SG parameters as presented here, the filter has been applied in several related contexts, primarily for smoothing NDVI time series and, less frequently, as a pre-classification spectral smoother. For instance, Kim et al. (2014) applied SG temporally to EVI and NDVI time series, rather than as a spatial or spectral pre-processing step for a single image, finding that temporal smoothing provided the best input for subsequent land cover classification. Similarly, Cao et al. (2018) and Yang et al. (2022) used temporal SG filtering to reconstruct cloud-contaminated time series data, consistently reporting improved results. However, these studies, like that of Kim et al., did not evaluate its use as a spectral pre-processing step for single-date image classification—the focus of our work. In a more distinct application, Roy (2022) developed a 2D shrinkage-based SG filter for geophysical potential field data, successfully estimating second derivatives even in the presence of significant noise, but likewise did not address the filter’s impact on classification tasks. General recommendations for data smoothing exist (e.g., Pandey, 2026), but systematic evaluations of SG parameters specifically for remote sensing image classification remain scarce.

Vaiphasa (2006) drew attention to the risks of subjectively chosen smoothing parameters, demonstrating that they can significantly alter the statistical properties of the input data. He showed that even a second-order SG filter can cause statistical disturbances (e.g., altering band separability as measured by a *t*-test) when used with excessively large kernel windows. Importantly, however, the kernel sizes tested by Vaiphasa (11, 31, 51) were considerably larger than the range we evaluated (3–19). Our preliminary analysis indicated that larger kernels

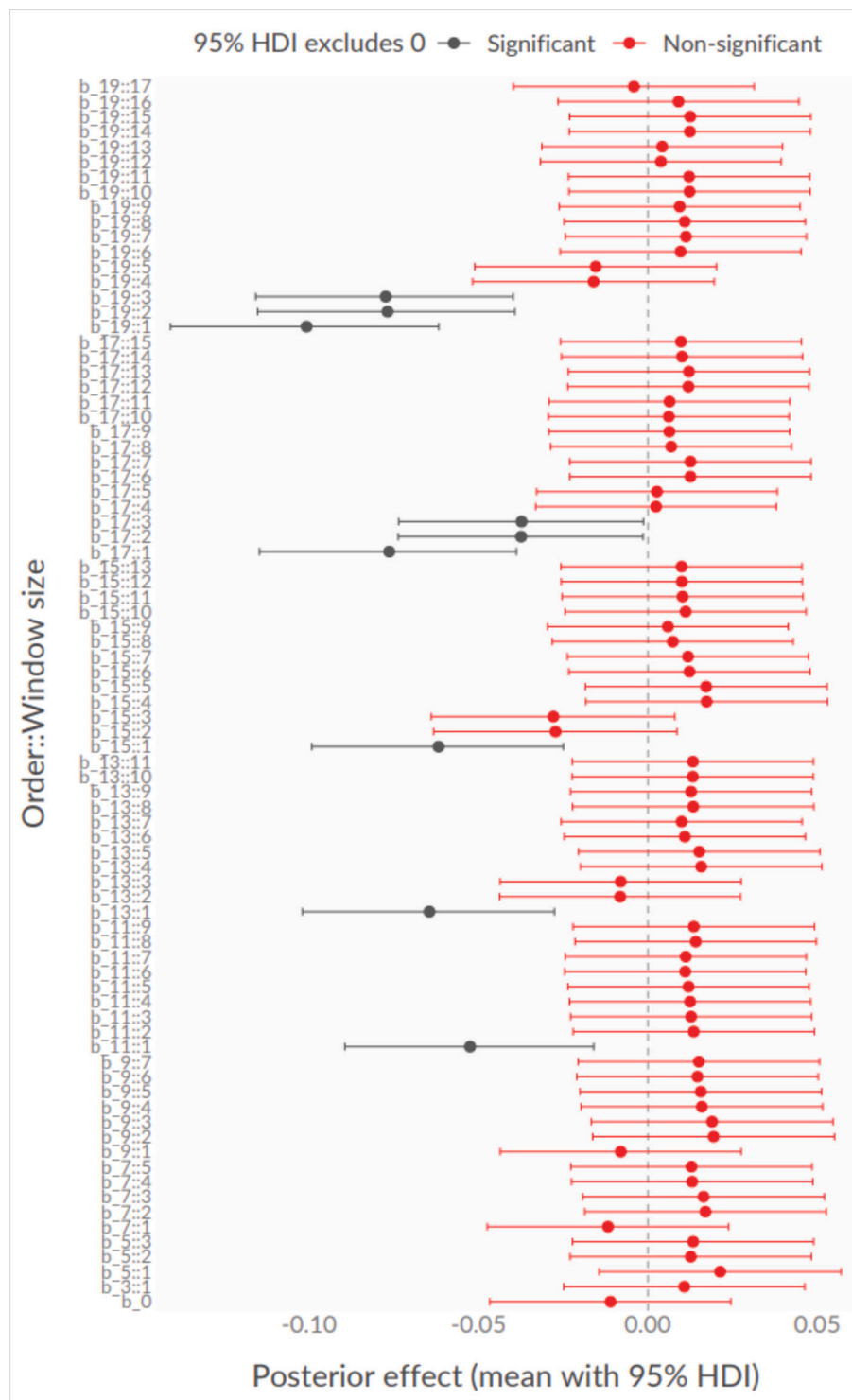


Fig. 10. Posterior estimates of the interaction between Window size and Order (denoted by::). Error bars represent the 95% highest density intervals (HDI) of the posterior distributions, with posterior means shown as dots. The vertical grey dashed line marks the null effect (zero). Nine interactions show significant effects (95% HDIs do not cross zero), while the remaining 72 parameter combinations (plus b_0 , the observed non-smoothed data) do not differ significantly from the overall mean.

led to unreasonable visual degradation of the image data, justifying our focus on a smaller parameter space.

The significance of only one interaction (Mecsek forest with pedunculate oak, Fig. 9) among twenty tested confirms that species influence is generally subordinate to, and mediated by, forest-level characteristics. The unique significance of pedunculate oak in the Mecsek site may be due to its particular spectral properties or abundance within that diverse canopy, making it more sensitive to the smoothing process.

The general non-significance of these interactions reinforces the conclusion that forest-level attributes outweigh the influence of individual species in determining the accuracy of different SG smoother parameterizations.

The observed improvements in classification accuracy, reaching absolute gains of up to 9%, demonstrate that spatial preprocessing can substantially influence hyperspectral forest classification performance. Recent studies have increasingly relied on computationally intensive

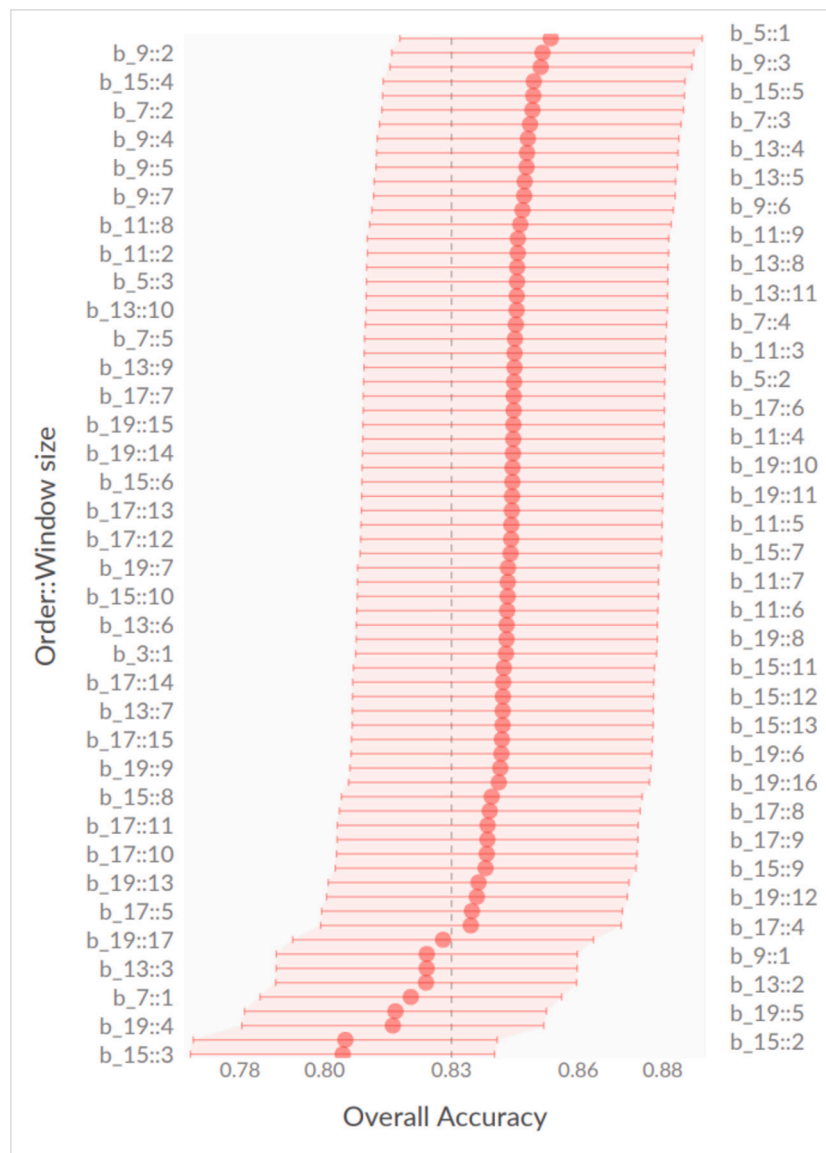


Fig. 11. Posterior estimates for the non-significant Window size and Order interactions (denoted by::), corresponding to the non-significant parameter combinations in Fig. 9. For each term (ordered top-to-bottom), transparent red error bars show the 95% HDI, with posterior means shown as transparent red dots. The vertical grey dashed line indicates the null effect (zero), defined as the mean (0.83) of the cross-validated, non-smoothed data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

deep learning architectures and advanced optimisation frameworks to address the high dimensionality and structural complexity of hyperspectral ecological data (Lin et al., 2024; Song and Wang, 2023; Veras and Ferreira, 2022). For example, Song and Wang (2023) combined Bayesian optimisation with feature selection techniques to improve leaf-level species classification from hyperspectral data, while SG filtering was employed only as an initial preprocessing step. In contrast, the present study specifically investigates the ecological and methodological behavior of SG smoothing, applying a Bayesian Hierarchical Random Model (BHRM) as the analytical framework to isolate the independent contributions of ecological and methodological factors. The results indicate that careful optimisation of spatial preprocessing can enhance crown-level spectral consistency while preserving relevant variability prior to classification. These findings suggest that statistically grounded preprocessing can complement advanced classification frameworks by improving the performance of ensemble approaches, such as Random Forest, in heterogeneous forest environments.

Although the BHRM framework employed here considers variability

among forest sites and species, the present study does not explicitly model fine-scale heterogeneity in stand structure (e.g., pure versus complex mixed stands). The degree of local species mixture and canopy structural complexity may influence the optimal level of spatial smoothing and classification performance. Incorporating spatially explicit stand composition metrics or plot-level structural inventory data into the hierarchical framework represents a promising direction for future research.

4.3. Methodological flexibility and robust smoother identification

A key methodological finding is the non-effectiveness of the majority (72 out of 81) of Order-Window size combinations (Fig. 11). This indicates substantial flexibility in parameter selection, as a wide range of configurations yield statistically equivalent accuracy. Among these non-effective smoothers, the combination of Window size 5 and 1 Order (b_5::1) was identified as the most robust due to its highest overall accuracy (Fig. 11). This provides a reliable, high-performance default

configuration for practical applications in heterogeneous forests, offering users a defensible starting point without the need for exhaustive parameter tuning.

It is important to note that the OA values analyzed in this study were not arbitrary accuracy metrics, but rather the product of a consistent Random Forest classification pipeline applied to both SG-filtered and non-smoothed images. This design ensures that the observed differences in OA are attributable to the pre-processing step (SG filtering) and the ecological context (*Forest, Species, Forest::Species*), rather than to variation in classifier configuration. The BHRM therefore isolates the independent contributions of ecological and methodological factors to classification accuracy, allowing us to quantify how preprocessing choices interact with ecological context to influence performance.

5. Conclusions

We aimed to disentangle the effects of study sites (encompassing image properties, phenophase, and climatic conditions) and species composition on the accuracy of species-level classifications from hyperspectral imagery processed with Savitzky-Golay (SG) filtering. By testing 81 SG parameter combinations (Order and Window size) across three structurally distinct Hungarian forests, we identified configurations that optimize classification accuracy while quantifying the relative importance of ecological and methodological factors.

Our analysis yielded three principal findings:

- 1. Forest structure overrides species identity.** The accuracy of SG smoothing is influenced more by forest-level structural and community characteristics than by the specific traits of individual tree species. This was evidenced by the significant effects observed for all three forest sites, while individual species showed no significant effects, and only one forest-species interaction (*Mecsek::pedunculata oak*) proved significant.
- 2. Substantial parameter flexibility exists.** The SG smoother's performance was robust across a wide range of parameter settings, with 72 of 81 Window size-Order combinations yielding statistically equivalent accuracies. This flexibility offers practical advantages, as users need not exhaustively tune parameters to achieve near-optimal results.
- 3. A practically robust default is available.** For applications requiring a single, stable configuration across diverse forests, the parameter combination of Window size 5 and Order 1 (b_5::1) is recommended based on its consistently high performance. However, the statistical equivalence of the 72 non-effective smoothers suggests that designating a single 'most robust' combination, while practically useful, is conceptually nuanced.

Beyond these specific findings, this study demonstrates the value of

adopting a hierarchical modelling framework when evaluating pre-processing techniques in remote sensing. The Bayesian Hierarchical Random Model (BHRM) proved well-suited to our data structure –where observations were nested within species and forests, and where we needed to account for both main effects and interactions simultaneously. By treating forest site, species, their interaction, and SG parameters as random effects, we were able to isolate their independent contributions to classification accuracy while acknowledging the inherent hierarchical structure of the data.

Critically, the BHRM framework is not limited to the specific conditions of this study. Because it handles complex, nested interactions and allows for the calibration of weakly informative priors, it can be deployed on hyperspectral datasets from different geographical regions, forest structures, or sensor specifications. Future researchers can adapt this framework to rigorously isolate optimal preprocessing parameters for their unique ecological contexts –whether evaluating SG filtering, alternative smoothing techniques, or other pre-processing methods altogether.

In summary, this study provides both practical guidance for SG parameter selection in heterogeneous forests and a flexible analytical template for evaluating pre-processing techniques in remote sensing. The combination of a comprehensive SG parameter assessment with a hierarchical Bayesian approach reveals that forest-level context governs smoothing performance, while offering the practical reassurance that a wide range of parameter choices –including the robust default b_5::1 –yield statistically equivalent results.

CRediT authorship contribution statement

Dávid Abriha: Writing – original draft, Visualization, Validation, Methodology, Data curation, Conceptualization. **Gergő Diószegi:** Writing – original draft, Visualization, Validation, Investigation, Conceptualization, Methodology. **Lilla Kovács:** Validation, Investigation. **Péter Burai:** Resources, Data curation. **Szilárd Szabó:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research was further supported by the National Research, Development and Innovation Office (NKFI Office) under grants NKFI K 138079 and NKFI ADVANCED 152950.

Appendix A

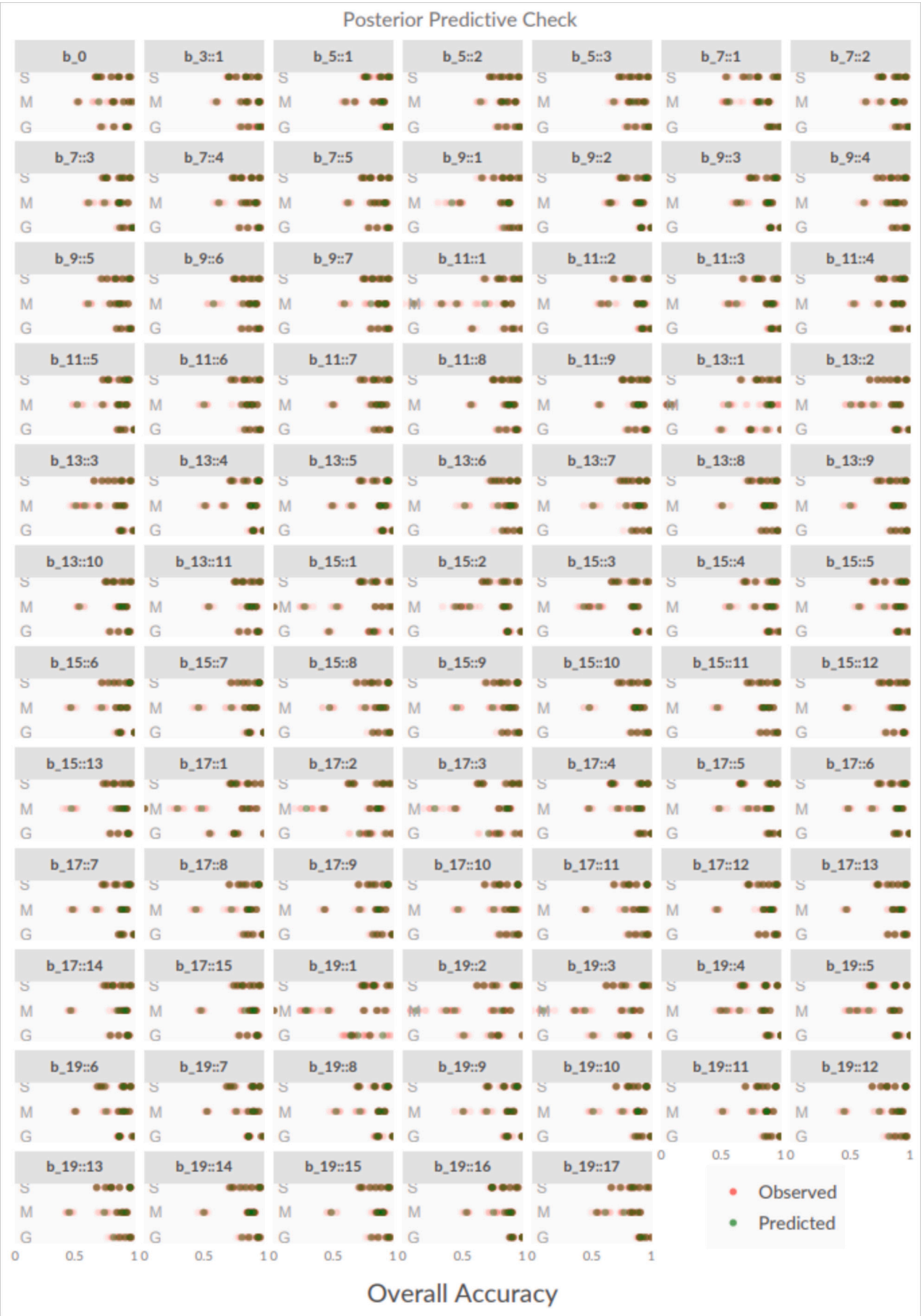


Fig. 12. Posterior predictive check for the Bayesian hierarchical model. The observed (non-smoothed) data points are shown in red, overlaid with the model's posterior predictive distribution (green points).

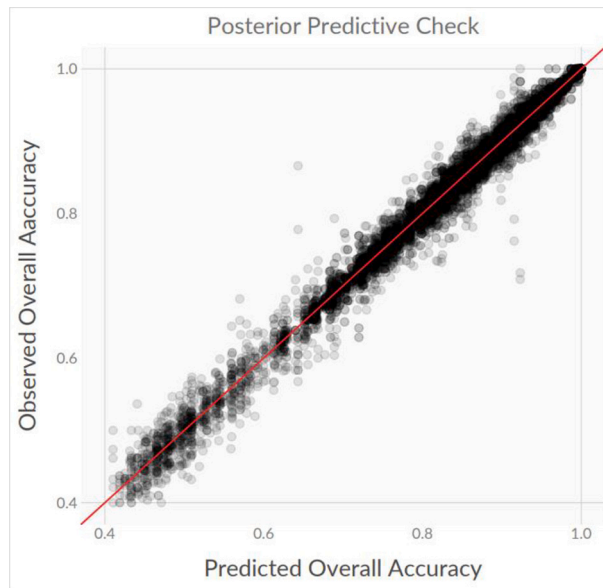


Fig. 13. Posterior predictive check: predicted versus observed overall accuracy. Points represent individual predictions, and the diagonal line indicates perfect agreement (1:1 relationship).

Data availability

Scripts and data are available under https://github.com/DijoG/SavGol_Smoothing.

References

- Abriha, Dávid, Srivastava, Prashant K., Szabó, Szilárd, 2023. Smaller is better? Unduly Nice accuracy assessments in roof detection using remote sensing data with machine learning and k-fold cross-validation. *Heliyon* 9 (3), e14045. <https://doi.org/10.1016/j.heliyon.2023.e14045>.
- Boyer, M., Miller, J., Belanger, M., Hare, E., Wu, Jiyou, 1988. Senescence and spectral reflectance in leaves of northern pin oak (*Quercus palustris* Muenchh.). *Remote Sens. Environ.* 25 (1), 71–87. [https://doi.org/10.1016/0034-4257\(88\)90042-9](https://doi.org/10.1016/0034-4257(88)90042-9).
- Burai, Peter, Laposi, Reka, Enyedi, Peter, Schmotzer, Andras, Kozma Bognar, Veronika, 2011. Mapping invasive vegetation using AISA eagle airborne hyperspectral imagery in the Mid-Ipoly-Valley. In: 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), June, pp. 1–4. <https://doi.org/10.1109/WHISPERS.2011.6080947>.
- Camps-Valls, G., Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 43 (6), 1351–1362. <https://doi.org/10.1109/TGRS.2005.846154>.
- Cao, Ruyin, Chen, Yang, Shen, Miaogen, et al., 2018. A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky-Golay filter. *Remote Sens. Environ.* 217 (November), 244–257. <https://doi.org/10.1016/j.rse.2018.08.022>.
- Chen, Jin, Jönsson, Per., Tamura, Masayuki, Gu, Zhihui, Matsushita, Bunkei, Eklundh, Lars, 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sens. Environ.* 91 (3), 332–344. <https://doi.org/10.1016/j.rse.2004.03.014>.
- Cho, Moses Azong, Debba, Praveesh, Mathieu, Renaud, Naidoo, Laven, van Aardt, Jan, Asner, Gregory P., 2010. Improving discrimination of savanna tree species through a multiple-endmember spectral angle mapper approach: canopy-level analysis. *IEEE Trans. Geosci. Remote Sens.* 48 (11), 4133–4142. <https://doi.org/10.1109/TGRS.2010.2058579>.
- Cutler, D. Richard, Edwards Jr., Thomas C., Beard, Karen H., et al., 2007. Random forests for classification in ecology. *Ecology* 88 (11), 2783–2792. <https://doi.org/10.1890/07-0539.1>.
- Dalponte, Michele, Örka, Hans Ole, Gobakken, Terje, Gianelle, Damiano, Næsset, Erik, 2013. Tree species classification in boreal forests with hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* 51 (5), 2632–2645. <https://doi.org/10.1109/TGRS.2012.2216272>.
- Dyderski, Marcin K., Paż, Sonia, Frelich, Lee E., Jagodziński, Andrzej M., 2018. How much does climate change threaten European Forest tree species distributions? *Glob. Chang. Biol.* 24 (3), 1150–1163. <https://doi.org/10.1111/gcb.13925>.
- Faraway, Julian, 2016. Extending the linear model with R: generalized linear. In: *Mixed Effects and Nonparametric Regression Models*, Second edition. <https://doi.org/10.1201/9781315382722>.
- Fassnacht, Fabian E., Neumann, Carsten, Förster, Michael, et al., 2014. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.* 7 (6), 2547–2561. <https://doi.org/10.1109/JSTARS.2014.2329390>.
- Gelman, Andrew, Carlin, John B., Stern, Hal S., Dunson, David B., Vehtari, Aki, Rubin, Donald B., 2013. *Bayesian Data Analysis*, 3rd ed. Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>.
- Ghiyamat, Azadeh, Shafri, Helmi Zulhaidi M., Mahdiraji, Ghafour Amouzad, Shariff, Abdul Rashid M., Mansor, Shattri, 2013. Hyperspectral discrimination of tree species with different classifications using single- and multiple-endmember. *Int. J. Appl. Earth Obs. Geoinf.* 23, 177–191. <https://doi.org/10.1016/j.jag.2013.01.004>.
- Gomez-Rubio, Virgilio, 2020. *Bayesian Inference with INLA*. CRC Press.
- Hirayama, Hidetake, Sharma, Ram C., Tomita, Mizuki, Hara, Keitarou, 2019. Evaluating multiple classifier system for the reduction of salt-and-pepper noise in the classification of very-high-resolution satellite images. *Int. J. Remote Sens.* 40 (7), 2542–2557. <https://doi.org/10.1080/01431161.2018.1528400>.
- Hycza, Tomasz, Stereńczak, Krzysztof, Balaży, Radomir, 2018. Potential use of hyperspectral data to classify forest tree species. *N. Z. J. For. Sci.* 48 (1), 18. <https://doi.org/10.1186/s40490-018-0123-9>.
- Immitzer, Markus, Atzberger, Clement, Koukal, Tatjana, 2012. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sens.* 4 (9), 9. <https://doi.org/10.3390/rs4092661>.
- Jia, Wen, Pang, Yong, Tortini, Riccardo, 2024. The influence of BRDF effects and representativeness of training data on tree species classification using multi-flightline airborne hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* 207 (January), 245–263. <https://doi.org/10.1016/j.isprsjprs.2023.11.025>.
- Jinhu, Bian, Li, Ainong, Mengqiang, Song, Liqun, Ma, Jiang, Jingang, 2010. Reconstruction of NDVI time-series datasets of MODIS based on Savitzky-Golay filter. *J. Remote Sens.* 14 (January).
- Kawala-Sterniuk, Aleksandra, Podpora, Michal, Pelc, Mariusz, et al., 2020. Comparison of smoothing filters in analysis of EEG data for the medical diagnostics purposes. *Sensors* 20 (3), 807. <https://doi.org/10.3390/s20030807>.
- Kim, So-Ra, Prasad, Anup, el-Askary, Hesham, et al., 2014. Application of the Savitzky-Golay filter to land cover classification using temporal MODIS vegetation indices. *Biol. Chem. Environ. Sci. Faculty Articles Res.* <https://doi.org/10.1016/j.jcis.2014.11.033> ahead of print, July 1.
- Kruschke, John, 2014. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.
- Kurz, A. Solomon, 2021. *Statistical Rethinking with Brms, Ggplot2, and the Tidyverse: Second Edition*. Version 0.2.0. <https://bookdown.org/content/4857/>.
- Li, Xian, Ding, Mingli, Pizurica, Aleksandra, 2020. Deep feature fusion via two-stream convolutional neural network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 58 (4), 2615–2629. <https://doi.org/10.1109/TGRS.2019.2952758>.

- Li, Xian, Gu, Yanfeng, Pizurica, Aleksandra, 2022. A unified multiview spectral feature learning framework for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60, 5540614. <https://doi.org/10.1109/TGRS.2022.3213838>.
- Li, Xian, Ding, Mingli, Gu, Yanfeng, Pizurica, Aleksandra, 2023. An end-to-end framework for joint denoising and classification of hyperspectral images. *IEEE Trans. Neural Netw. Learn. Syst.* 34 (7), 3269–3283. <https://doi.org/10.1109/TNNLS.2023.3264587>.
- Liang, Hu, Li, Na, Zhao, Shengrong, 2021. Salt and pepper noise removal method based on a detail-aware filter. *Symmetry* 13 (3), 3. <https://doi.org/10.3390/sym13030515>.
- Likó, Szilárd Balázs, Bekő, László, Burai, Péter, Holb, Imre J., Szabó, Szilárd, 2022. Tree species composition mapping with dimension reduction and post-classification using very high-resolution hyperspectral imaging. *Sci. Rep.* 12 (1), 1. <https://doi.org/10.1038/s41598-022-25404-x>.
- Likó, Szilárd Balázs, Holb, Imre J., Oláh, Viktor, Burai, Péter, Szabó, Szilárd, 2024. Deep learning-based training data augmentation combined with post-classification improves the classification accuracy for dominant and scattered invasive forest tree species. *Remote Sens. Ecol. Conserv.* <https://doi.org/10.1002/rse2.365>.
- Lin, Feng-Cheng, Shiu, Yi-Shiang, Wang, Pei-Jung, Wang, Uen-Hao, Lai, Jhe-Syuan, Chuang, Yung-Chung, 2024. A model for forest type identification and forest regeneration monitoring based on deep learning and hyperspectral imagery. *Ecol. Inform.* 80 (May), 102507. <https://doi.org/10.1016/j.ecoinf.2024.102507>.
- Liu, Haijian, Wu, Changshan, 2018. Crown-level tree species classification from AISA hyperspectral imagery using an innovative pixel-weighting approach. *Int. J. Appl. Earth Obs. Geoinf.* 68 (June), 298–307. <https://doi.org/10.1016/j.jag.2017.12.001>.
- Lukin, Vladimir, Abramov, Sergey, Krivenko, Sergey, Kurekin, Andriy, Pogrebnyak, Oleksiy, 2013. Analysis of classification accuracy for pre-filtered multichannel remote sensing data. *Expert Syst. Appl.* 40 (16), 6400–6411. <https://doi.org/10.1016/j.eswa.2013.05.061>.
- Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* 42 (8), 1778–1790. <https://doi.org/10.1109/TGRS.2004.831865>.
- Modzelewska, Aneta, Fassnacht, Fabian Ewald, Stereńczak, Krzysztof, 2020. Tree species identification within an extensive forest area with diverse management regimes using airborne hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* 84 (February), 101960. <https://doi.org/10.1016/j.jag.2019.101960>.
- Nepi, Lindo, Quattrini, Giacomo, Pesaresi, Simone, Mancini, Adriano, Pierdicca, Roberto, 2025. AI-based estimation of forest plant community composition from UAV imagery. *Eco. Inform.* 90 (December), 103199. <https://doi.org/10.1016/j.ecoinf.2025.103199>.
- Pandey, Puneeta, 2026. Remote Sensing & GIS Applications in Environmental Science – Simple Book Publishing. Accessed February 17, 2026. <https://ebooks.inflibnet.ac.in/esp06/>.
- Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, et al., 2018. Scikit-learn: machine learning in Python. *arXiv*. <https://doi.org/10.48550/arXiv.1201.0490>, 1201.0490. Preprint, arXiv, June 5.
- Piironen, Rami, Fassnacht, Fabian Ewald, Heiskanen, Janne, Maeda, Eduardo, Mack, Benjamin, Pellikka, Petri, 2018. Invasive tree species detection in the eastern Arc Mountains biodiversity hotspot using one class classification. *Remote Sens. Environ.* 218 (December), 119–131. <https://doi.org/10.1016/j.rse.2018.09.018>.
- Qin, Haiming, Zhou, Weiqi, Yao, Yang, Wang, Weimin, 2022. Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data. *Remote Sens. Environ.* 280 (October), 113143. <https://doi.org/10.1016/j.rse.2022.113143>.
- R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richter, Ronny, Reu, Björn, Wirth, Christian, Doktor, Daniel, Vohland, Michael, 2016. The use of airborne hyperspectral data for tree species classification in a species-rich central European forest area. *Int. J. Appl. Earth Obs. Geoinf.* 52 (October), 464–474. <https://doi.org/10.1016/j.jag.2016.07.018>.
- Roy, Indrajit G., 2022. Regularized 2D Savitzky–Golay derivative filter in estimating the second vertical derivative of potential field data. *Geophys. Prospect.* 70 (6), 1066–1081. <https://doi.org/10.1111/1365-2478.13210>.
- Rue, Håvard, Martino, Sara, Chopin, Nicolas, 2009. Approximate Bayesian inference for latent gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat Methodol.* 71 (2), 2. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>.
- Ruffin, Chris, King, Roger L., Younan, Nicolas H., 2008. A combined derivative spectroscopy and Savitzky–Golay filtering method for the analysis of hyperspectral data. *GISci. Remote Sens.* 45 (1), 1–15. <https://doi.org/10.2747/1548-1603.45.1.1>.
- Savitzky, Abraham, Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36 (8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- Shang, Xiao, Chisholm, Laurie A., 2014. Classification of Australian native Forest species using hyperspectral remote sensing and machine-learning classification algorithms. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.* 7 (6), 2481–2489. <https://doi.org/10.1109/JSTARS.2013.2282166>.
- Shebl, Ali, Abriha, Dávid, Fahil, Amr S., El-Dokouny, Hanna A., Elrasheed, Abdelmajeed A., Csámer, Árpád, 2023. PRISMA hyperspectral data for lithological mapping in the Egyptian Eastern Desert: evaluating the support vector machine, random forest, and XG boost machine learning algorithms. *Ore Geol. Rev.* 161 (October), 105652. <https://doi.org/10.1016/j.oregeorev.2023.105652>.
- Shi, Yifang, Skidmore, Andrew K., Wang, Tiejun, et al., 2018. Tree species classification using plant functional traits from LiDAR and hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* 73 (December), 207–219. <https://doi.org/10.1016/j.jag.2018.06.018>.
- Song, Guangman, Wang, Quan, 2023. Species classification from hyperspectral leaf information using machine learning approaches. *Eco. Inform.* 76 (September), 102141. <https://doi.org/10.1016/j.ecoinf.2023.102141>.
- Vaiphava, Chaichoke, 2006. Consideration of smoothing techniques for hyperspectral remote sensing. *ISPRS J. Photogramm. Remote Sens.* 60 (2), 91–99. <https://doi.org/10.1016/j.isprsjprs.2005.11.002>.
- Van Houtven, George, Phelan, Jennifer, Clark, Christopher, et al., 2019. Nitrogen deposition and climate change effects on tree species composition and ecosystem services for a forest cohort. *Ecol. Monogr.* 89 (2), e01345. <https://doi.org/10.1002/ecm.1345>.
- Vargas-Muñoz, John E., Tuia, Devis, dos Santos, Jeferson A., Falcão, Alexandre X., 2017. Post classification smoothing in sub-decimeter resolution images with semi-supervised label propagation. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July, pp. 3688–3691. <https://doi.org/10.1109/IGARSS.2017.8127799>.
- Veras, Hudson Franklin Pessoa, Ferreira, Matheus Pinheiro, da Cunha Neto, Ernandes Macedo, Figueiredo, Evandro Orfanó, Corte, Ana Paula Dalla, Sanquetta, Carlos Roberto, 2022. Fusing multi-season UAS images with convolutional neural networks to map tree species in Amazonian forests. *Eco. Inform.* 71 (November), 101815. <https://doi.org/10.1016/j.ecoinf.2022.101815>.
- Woodall, C.W., Oswalt, C.M., Westfall, J.A., Perry, C.H., Nelson, M.D., Finley, A.O., 2010. Selecting tree species for testing climate change migration hypotheses using forest inventory data. *For. Ecol. Manag. Adapt. For. Forest Manag. Chang. Clim.* 259 (4), 778–785. <https://doi.org/10.1016/j.foreco.2009.07.022>.
- Yang, Xue, Chen, Jin, Guan, Qingfeng, Gao, Huan, Xia, Wei, 2022. Enhanced spatial-temporal Savitzky–Golay method for reconstructing high-quality NDVI time series: reduced sensitivity to quality flags and improved computational efficiency. *IEEE Trans. Geosci. Remote Sens.* 60, 1–17. <https://doi.org/10.1109/TGRS.2022.3190475>.
- Yulianto, Fajar, Nugroho, Gatot, Chulafak, Galdita Aruba, Suwarsono, Suwarsono, 2021. Improvement in the accuracy of the postclassification of land use and land cover using Landsat 8 data based on the majority of segment-based filtering approach. *Sci. World J.* 2021 (March), 6658818. <https://doi.org/10.1155/2021/6658818>.
- Zhang, Jiawei, Fu, Pingjie, Meng, Fei, Yang, Xinyue, Xu, Jie, Cui, Yu, 2022. Estimation algorithm for chlorophyll-a concentrations in water from hyperspectral images based on feature derivation and ensemble learning. *Eco. Inform.* 71 (November), 101783. <https://doi.org/10.1016/j.ecoinf.2022.101783>.
- Zhao, Hengwei, Zhong, Yanfei, Wang, Xinyu, et al., 2022. Mapping the distribution of invasive tree species using deep one-class classification in the tropical montane landscape of Kenya. *ISPRS J. Photogramm. Remote Sens.* 187 (May), 328–344. <https://doi.org/10.1016/j.isprsjprs.2022.03.005>.
- ZhiYong, Lv, Shi, WenZhong, Benediktsson, Jón Atli, Gao, LiPeng, 2018. A modified mean filter for improving the classification performance of very high-resolution remote-sensing imagery. *Int. J. Remote Sens.* 39 (3), 770–785. <https://doi.org/10.1080/01431161.2017.1390275>.
- Zhu, Xuan, 2013. Land cover classification using moderate resolution satellite imagery and random forests with post-hoc smoothing. *J. Spat. Sci.* 58 (2), 323–337. <https://doi.org/10.1080/14498596.2013.819600>.