Clustering of hydrological variables across a coastal forest bordering a salt marsh

Giovanna Nordio, Keryn Gedan, Sergio Fagherazzi

Abstract

Sea level rise and storm surges drive coastal forest retreat and salt marsh expansion. Salt intolerant tree species cannot survive flooding and high soil salinity. Salinization is exacerbated by long-lasting flooding events. Both salinization and flooding drive ecological zonation in coastal areas. Hydrological variables, if coupled with ecological surveys, can explain the different stages of coastal forest retreat and marsh encroachment. In this research, a long-term series of hydrological variables collected along transects from marsh to inner forest are analyzed. Linear discriminant analysis (LDA) is used to identify the hydrological variables responsible for the forest-marsh gradient and their seasonal patterns. Water content in the soil (WC) and groundwater electrical conductivity (EC) were found to be the main variables responsible for the hydrological difference among the sites. Higher values of WC and EC were found in the low forest area near the marsh, as it is reflected by ecological data. Seasonal LDA data suggest that the sites are more hydrologically different during winter (higher distance among clusters of variables) and similar during summer (less distance among clusters). In the study area, higher rainfall occurred during summer, decreasing groundwater EC in areas characterized by less canopy cover (dying forest). Evapotranspiration in the low forest is higher due to more radiation reaching the soil and new invasive vegetation. Both processes move these sited closer to the high forest, which is less impacted by sea level rise. During storm surge events the distance between clusters also decreases, because of uniform salinization and flooding. We conclude that vegetation gradients at the forest-marsh boundary are supported by hydrological gradients during winter months in the absence of storm surges. An additional analysis is carried out to determine whether hydrological values

collected in each forested area are closer to those of the healthy forest or to the marsh. Results clearly support the previous LDA analysis and the ecological data collected, showing that negative conditions for trees occur closer to the marsh; these negative conditions are exacerbated during storm surge events also in the healthy forest.

Introduction

Saltwater intrusion and flooding triggered by SLR (Sea Level Rise) and storm surges are the main drivers of coastal forest retreat and consequent marsh expansion (Kearney et 2019; Schieder & Kirwan, 2019; Tully et al. 2019; Doyle et al. 2010; DeSantis et al. 2007; Fagherazzi et al. 2019). Ghost forests, consisting of dead tree stands adjacent to marshes, are clear indicators of coastal forest retreats (Schieder & Kirwan, 2019). This process is prominent in low-lying areas, where saltwater intrusion and flooding can affect the coastal groundwater from above and below (Michael et al. 2013; Yang et al. 2018; Gleeson et al. 2011). Both higher sea levels and local geomorphology make these areas more prone to coastal transgression.

Salinization and flooding events are threats for salt intolerant vegetation species and encourage ecological zonation. Shallow water table and high groundwater salinity affect tree survival, lowering photosynthetic activity (Antonellini & Mollema, 2010; Munns & Tester 2008). Increasing salinity values progressively kill salt-intolerant tree species, encouraging expansion of invasive vegetation and salt marsh establishment (Smith et al. 2013). For example, *Phragmites australis* is usually found at the lower border of a forest, where salinity conditions are brackish (Smith et al. 2013).

Rainfall and evapotranspiration also affect local hydrology. Rainfall events can have a positive effect on coastal aquifers, reducing groundwater and soil salinity (Summer & Belaineh, 2005;

Allen et al. 1998; Gardner et al., 1992). Vegetation canopy, radiation, temperature, wind intensity and pressure are the main variables regulating evapotranspiration (Cai et al. 2007). In a dense forest, evapotranspiration of understory vegetation is low, because of the limited radiation penetrating the canopy (Cao et al. 2006).

Groundwater dynamics along coastal areas are often studied with numerical models (ex. Paldor & Michael, 2021; Geng et al. 2020; Yu et al. 2016) and field data (ex. Gardner et al. 2002; Thibodeau et al. 2011). Models simulate the overall response of groundwater to external inputs as a function of hydro-geological features in idealized domains. Field data allow to quantify local groundwater dynamics, determining the response to external inputs such as precipitation, evapotranspiration and storm surges. Collecting and studying long-term datasets is time-consuming. Here we present a novel approach to analyze hydrological data at the marsh forest boundary.

Linear discriminant analysis (LDA) is mostly used to cluster and classify multivariate data (Rencher & Christensen, 2012). Ecologists and biologists used LDA to identify patterns in the collected datasets and the variables driving group separation (Stewart et al. 2013). For instance, ecological variations were estimated in the Peruvian Amazon using discriminant analysis (Correa-Metrio et al. 2010). Vegetation data were classified to determine suitable habitats for avian species (Rice et al. 1983); water quality and physical-chemical data were also correlated to water plant species incidence (Wiegleb 1981). As far as we know, LDA was never used to cluster multivariate hydrological datasets. In this paper we apply LDA on groundwater and soil data collected in a coastal forest bordering a salt marsh from 2019 to 2022. Through this technique, we can spatially visualize the hydrological separation between groups at different ecological stages and relate results to the ecological variables collected in fieldwork campaigns. The strength of LDA is that

we can fast know where hydrological conditions differ the most, which variables determine the difference, and how the group separation changes over time.

Overall, the main goal of our analysis is: (1) to reduce the hydrological multivariate space in a simple 2-D space; (2) to determine key variables controlling the hydrological separation among sites; (3) to correlate clustering based on hydrological variables to ecological parameters at each forest site; (4) to interpret seasonal variations of hydrological parameters affecting the groundwater and soil dynamics. This approach produces a first overall snapshot of local hydrology, speeding up the analysis of intensive long-term datasets.

Methods

Study area

We focus our research on a coastal forest bordering a salt marsh in Nassawadox, Delmarva Peninsula (VA) (Fig. 1a). The forest, occupying an area of 3 ha, is characterized by an average elevation of 1 m on NAVD88 (North America Vertical Datum) and a distance from the salt marsh ranging from 140 and 600 m. Soil samples collected every 20 cm until 1m below the ground surface classified the area as averagely consisting of 80% clay, 10% sand and 10% vegetation/roots. The forest is dominated by loblolly pines (*Pinus taeda*), and it is more or less affected by saltwater intrusion and flooding depending on distance from the salt marsh. Close to the salt marsh, (low forest) trees are barren or partially dead, and *Phragmites australis* encroaches the area (Chambers et al. 1999). Moving inland, the percentage of healthy trees increases. At intermediate distances, 150 m from the marsh (medium forest), shrubs of *Myrica cerifera* coexist with relatively healthy trees. At the farthest distances from the forest-marsh boundary (high forest) healthy trees dominate. Six different forested sites were considered in this analysis. Two in the

high forest, H5 and H7, two in the medium forest M1 and M2 and two in the low forest L1 and L6 (Fig. 1b). The described ecological patterns represent the evolution of the forest, affected by sea level rise and storm surges (Fig. 1b). In this area the climate is humid subtropical, with temperatures ranging from 5 °C in winter and 28 °C in summer, accordingly to NOAA (National Oceanic and Atmospheric Administration). During the summer season, conspicuous precipitation events occur frequently, and rainfall amounts are higher. According to data collected between 1991 and 2020, mean precipitation measured at the NOAA Norfolk station (US USW00013750), reaches its monthly maximum of 13 cm in August, and rainy days are maximum in summer (around 15 days). The shoreline is characterized by a semidiurnal tide, with mean amplitude of 2 m on NAVD88 (NOAA station: 8631044 Wachapreague). In this area storm surges can occur during fall, mostly triggered to tropical cyclones; or during winter, caused by nor'easter storms. Nor'easters form at higher latitudes along the East Coast of North America, bringing wind, snow, rain and flooding to these regions.

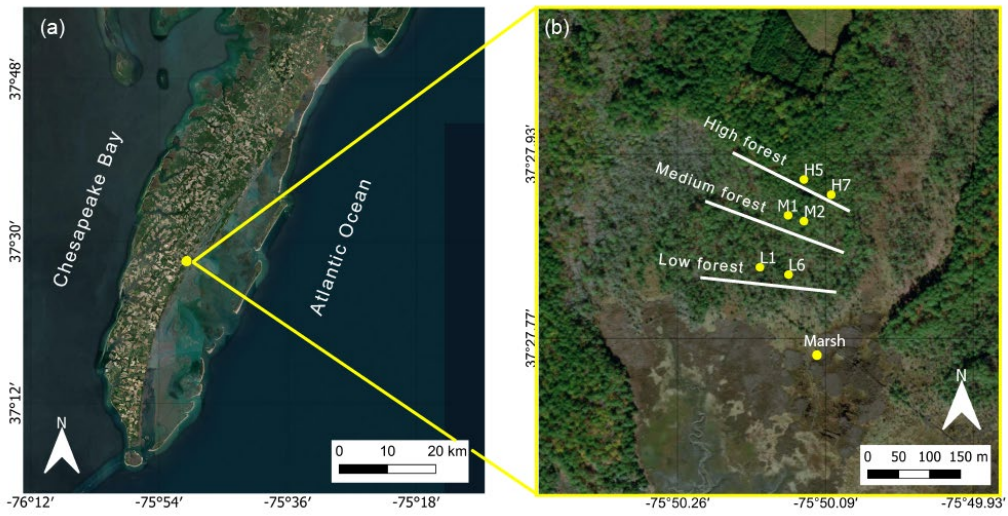

Figure 1: (a) Study area, (b) different ecological zonation and sites in the coastal forest. Images from ESRI (Environmental System Research Institute) Satellite Imagery.

Hydrological Data

Groundwater and soil moisture data were collected in six forested sites. Groundwater level, temperature and specific electrical conductivity (at 25 °C) were measured inside screened wells, 1m deep below the ground surface, using a CTD-diver (Van Essen). Soil water content, temperature, and specific electrical conductivity (at 25 °C) were measured by TEROS devices (Meter Group), at each site. The sensors were placed 7 cm below the ground surface. Data collection started in January 2019 and data were recorded every hour. Data can be found in the VCR-LTER (Virginia Coast Reserve - Long Term Ecological Research) data portal (dataset ID:VCR22344) (Fagherazzi & Nordio, 2021). Here we consider data from June 2019 to May 2022. During this timeframe, four main storm surge events affected groundwater at all sites: Melissa tropical storm, which occurred between October 11-14, 2019, when sea level reached 1.42 m on NAVD88 at the Wachapreague NOAA station (id:8631044); two storm surges also occurred in May 30 and October 10, 2021, when sea level respectively reached 1.39 m and 1.37 m on NAVD88. Finally, in January 3, 2022 a strong storm surge was felt in each site, increasing sea level to 1.55 m on NAVD88. Tropical storm Wanda, which occurred between October 26 and November 7, 2021, was partially felt by the sensors, since groundwater level and electrical conductivity were already high from the previous storm surge. Four storm surges occurred in August 2019, April 2020,September 2020 and November 2021 and only affected the low forested sites. Precipitation data collected by NOAA in Norfolk city (VA) and in Salisbury (MD) (https://www.ncdc.noaa.gov/cag/) are used to correlate seasonal groundwater and soil data to freshwater external inputs.

Hydrological data are coupled with ecological data collected at the same sites and reported in the VCR-LTER portal dataset (ID: VCR22354) (Gedan et al. 2022). Tree species were first identified

and counted inside a 20x20m plot. For each tree, diameter at breast height (DBH) was measured at 1.37 m from tree base. Basal area, considered the most significant variable describing the difference between forest stages, was consequently calculated. An average basal area, independent of tree species, is considered in future analysis. Shrub species were identified in a 2.5 m-radius circle in each plot. We concentrate on the dominant species *Myrica cerifera*. Again, shrub basal area was calculated and analyzed. Non-woody vegetation data were collected in four to eight randomly placed 1x1m plots at each site. For each species, the percent cover was estimated using Domin - Krajina scale (Mueller-Dombois & Ellenberg, 1974). We concentrate the analysis on *Phragmites australis*, the species that firstly invade stressed forests and precedes marsh expansion. In fact, *Phragmites australis* can survive in brackish conditions (Smith et al. 2013).

Linear Discriminant Analysis

Discriminant analysis (DA) is a powerful tool to cluster and classify multivariate observations (Balakrishnama, S., & Ganapathiraju, A. 1998; Xanthopoulos et al. 2013). This analysis is used to examine the group separation in a two-dimensional space, where points of an original *p*-dimensional space are represented in the best possible view, reducing their dimensionality. DA also determines the subset of original variables best separating the groups, and allows to rank the variables in terms of their contribution to the groups' separation. Finally, the obtained discriminant functions can be used to allocate new points (Rencher and Christensen, 2012; Izenman 2013). Discriminant analysis supervised classifier is one of the basic and simplest classifiers used in many clustering and classification problems. There are two main types of DA classifiers, the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) classifiers (Tharwat 2016). When the number of variables in the dataset is much higher than the number of samples for each class, regularized discriminant functions are preferred (Tharwat 2016; Wu et al. 1996).

Before conducting a DA analysis, we need to check that the data conform to a multivariate normal distribution. Then, given the original variables Y_i (i=1,…n), a linear combination of the original variables can be generated to separate groups :

$$Z_i = a_{i1}Y_1 + a_{i2}Y_2 + \cdots + a_{in}Y_n \quad (1)$$

The goal is to find the coefficients **a** that maximize the difference among groups. In other words, we need to find the largest among-group sums of squares for a given set of Z_i's (Rencher and Christensen, 2012; Gotelli and Ellison 2004). Considering H as the among-groups sum of squares and cross-products (SSCP) matrix and E as the between-groups SSCP, coefficients $\boldsymbol{a}$ can be found solving:

$$\boldsymbol{a}'(H\boldsymbol{a} - \lambda E\boldsymbol{a}) = 0 \quad (2)$$

After determining the eigenvalues λ, the eigenvectors **a** can be finally calculated (Rencher and Christensen, 2012). Eigenvalues measure the importance of each discriminant function Z_i. Wilks' Λ-test, with χ-approximation, is finally conducted on the eigenvalues to assess the significant *p* dimensions of group separation (Rencher and Christensen, 2012). Principal component analysis (PCA) is a similar classifier, using linear combination of variables as previously described. Differently from DA, PCA is an unsupervised method that find the greatest extent of variance in a set of data (Stewart et al. 2014).

Our dataset consists of 6 original variables: groundwater level (WL), groundwater conductivity (EC), groundwater temperature (T), soil water content (WC), soil conductivity (SEC) and soil temperature (ST). Variables are measured in each forest site. First, we test the normality distribution assumption. Second, we check for correlation among variables (linear correlation with significance level of 95%). This is important because multicollinearity can produce misleading

results. If two variables are strongly correlated, one of them is removed from the final analysis. Because the variables describe different processes and have different unit of measurements, they are standardized (mean μ_z=0, variance σ=1).

We then conduct yearly and seasonal DAs on the data. All the linear discriminant classifiers are first trained using 60% of data and then tested using the remaining 40% of data. Euclidean distances between centroids are compared over the seasons and years using ANOVA test and post-hoc Tukey test with 90% significance level. `lda` command from `MASS` library in R-studio is used to conduct the analyses. Finally, an interpretation of results is provided on the basis of ecological and hydro-meteorological data. A similar approach, using linear functions, allows us to reduce data dimensionality and to define the most dominant variables responsible for the ecological spatial patterns, giving an overview of the differences between sites and the effects of external hydrological inputs.

Finally, an additional analysis is proposed. Data collected in H5 and in the marsh from June 2021 to May 2022 are firstly classified using linear discriminant analysis. H5 data are considered as representative of 'Healthy forest' group and marsh data of 'Marsh' group. Data from the remaining groups are then classified based on the linear model obtained. According to Rencher and Christensen (2012), in a two-group classification an observation is assigned to 'Healthy forest' (group 1) if:

$$Z = \boldsymbol{a}'Y > \frac{1}{2}(\overline{Z_1} + \overline{Z_2}) \quad (3)$$

where $\overline{Z_1}$ and $\overline{Z_2}$ are the centroids of each group. Otherwise, it is assigned to the 'Marsh' group (group 2). In this way, we can estimate when and where forest is more stressed and compare the results with the ecological data.

Results

Hydrological data variations over the seasons

Real data of groundwater level, conductivity, temperature, and soil water content, conductivity and temperature are shown in Fig. 2. According to the post-hoc Tukey test results, mean value of variables is significantly different ($\alpha<0.05$) over the seasons. Groundwater level and conductivity, water content and soil conductivity (Fig. 2a,b,d,e) are usually lower during the summer season. Groundwater and soil temperature reach their peaks during summer season (Fig. 2c,f). Storm surge effects are mostly felt as increases in groundwater conductivity (fall 2019, fall 2021) (Fig. 2b). From Fig. 2 and from the datasets reported in VCR-LTER portal (Fagherazzi & Nordio, 2021), we can see the effects of the external inputs on the hydrological variables, but we struggle to easily explain the overall hydrological difference among the forested areas.

Forest structure

Mean tree basal areas measured in the high forest (H5 and H7) are significantly higher than those measured in low forest and intermediate forest (*p-value* <0.1) (Fig. 3a). This confirms the healthy state of trees in the high forest. No significant difference can be found in the shrub basal areas, in plots where shrubs of *Myrica cerifera* are present (Fig. 3b). Phragmites australis is present in each site except H5. The percent coverage is significantly higher in H7, M2 and L1 compared to the other stations (Fig. 3c).

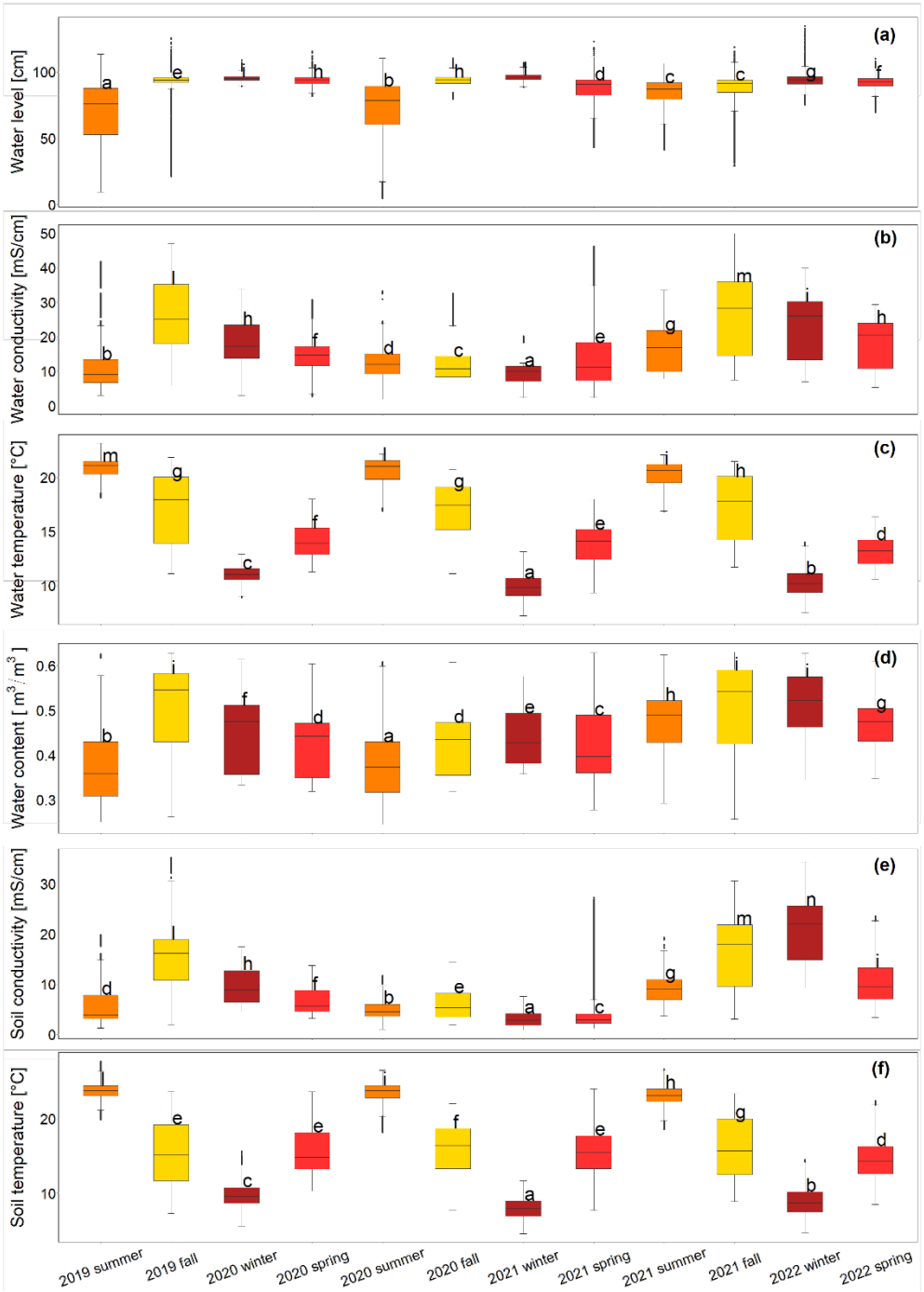

Figure 2: Real variables distribution over the season. Datasets are reported in the VCR-LTER data portal (dataset ID:VCR22344) (Fagherazzi & Nordio, 2021).

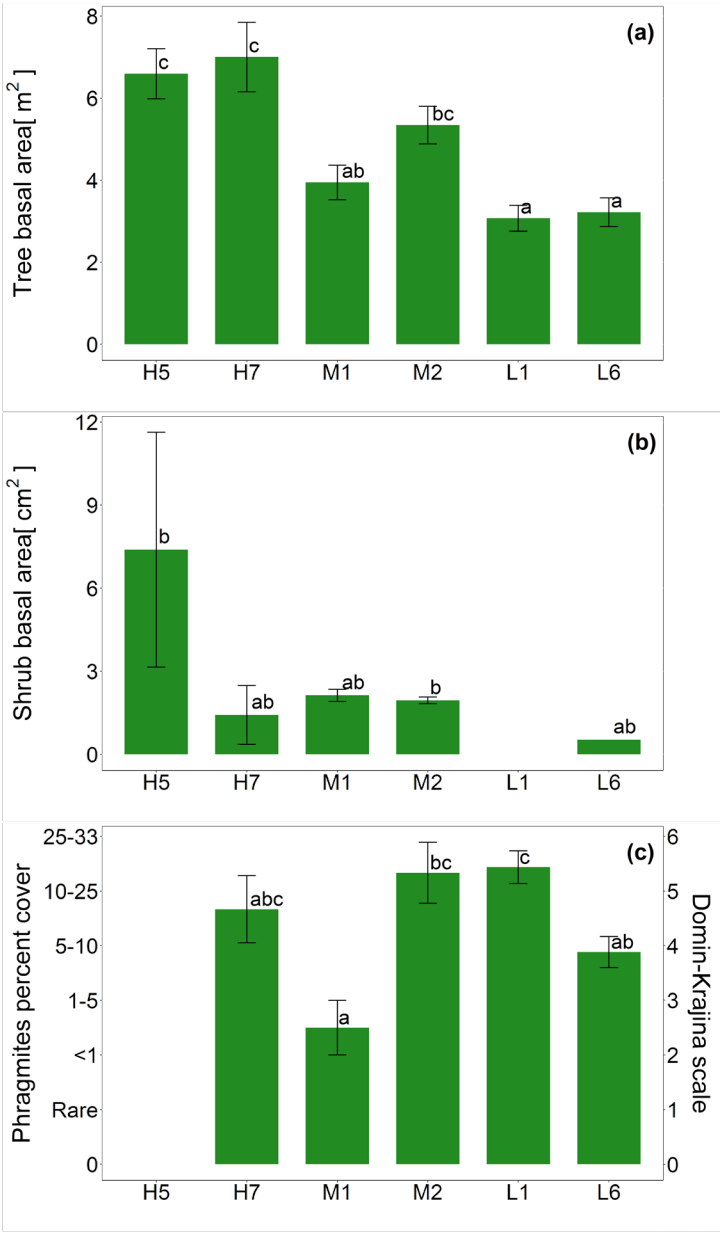

Figure 3: (a) Tree basal area and (b) shrub basal area in six forested sites. (c) Praghmatis australis percent coverage and correspondent Domin-Krajina scale index in the three forested areas. Letters identify the post-hoc Tukey test results with 90% confidence level (Gedan et al. 2021).

LDA yearly analysis and cluster identification

LDA results, conducted on yearly data collected in 2020 and 2021, are shown in Fig. 2. In 2020 no storm surge events were detected, while in 2021 two main storm surges occurred. Since ST and T were significantly correlated (R^2=0.88, $p<0.05$), ST is neglected in the LDA. The first two linear discriminant functions LD1 and LD2 can explain more than 90% of the group distribution in a 2D-plane. Accuracy of training and testing data model are around 96% for 2020 dataset and 80% for 2021 dataset (Table 1-2, in Supplementary material). Water content in the soil (WC) and groundwater electrical conductivity (EC) are the dominant variables in LD1 and LD2, respectively, in for groups' separation (Table 3 in Supplementary material). In 2021, the contribution of the other variables increases. Linear discriminant classifiers can significantly ($\chi^2>\chi_{crit}^2$) separate groups in the 2020 dataset. LD1 and LD2 for high forested sites range respectively from -7 to -2 and from -5 to 5. This suggests low WC and EC values in the high-forest sites. Medium-forest values range from -2 to 5 in LD1 and from -2 to 2 in LD2, indicating modest WC and EC values in the medium site. Low forest sites differ from each other. LD2 values for L1 are negative, while they are positive for L6 (Fig. 4a). This suggests that the L1 cluster is dominated by high electrical conductivity while L6 by high soil water content (Fig. 4a). The group separation, expression of the hydrological differences, suggests not only a difference in the ecology of the three forested zones (low, medium, and high), but also an ecological difference between sites in the same zone, in the high and low forest. Ellipses, representing variance of data, are larger for clusters L1 and L6. LDA performed on the 2021 dataset does not clearly separate groups in a 2D-plane (only LD1 $\chi^2>\chi_{crit}^2$). EC and WC are still the dominant variables in the group separation and variances are quite large for almost all clusters (Fig. 4b). This cluster collapsing can be correlated to storm surge events occurring in spring and fall 2021. Centroids coordinates in both 2020 LDA and 2021 LDA are

calculated and compared (Fig. 5, Fig. S1 in Supplementary material). On average, distance between clusters in the 2020 analysis is significantly higher (p<0.1) than the distance in 2021. Moreover, in 2020 distances between sites in the same forest area are reduced and distances among sites in different areas are maximum (Fig. S1 in Supplementary material).

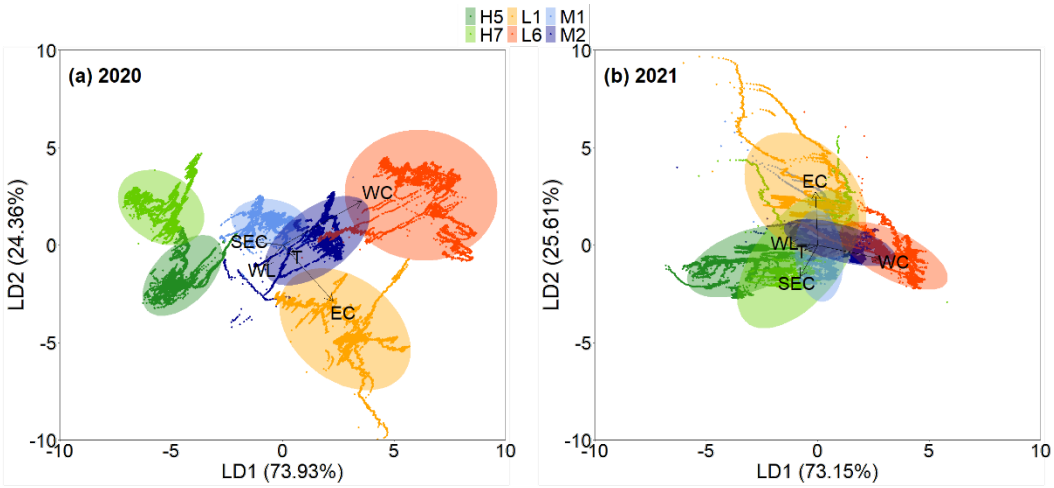

Figure 4: LDA results for datasets collected in 2020 (a) and 2021 (b). Arrows are proportional to the linear function loadings. WL=groundwater level, EC= groundwater electrical conductivity, T=temperature, WC= soil water content, SEC=soil electrical conductivity.

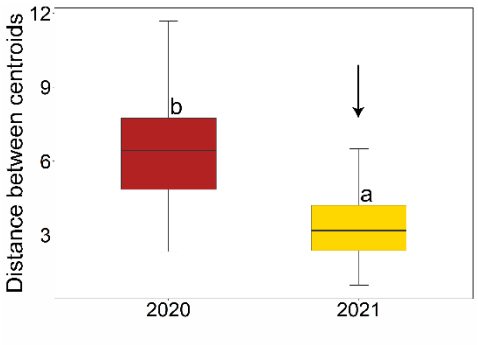

Figure 5: Euclidean distances between centroids in 2020 and 2021. Letters represent the post-hoc test results after t-test (90% significance level). Black arrow identifies the occurrence of storm surge events.

Seasonal patterns, external inputs and ecological change

LDA analysis for each season from summer 2019 to spring 2022 is conducted (Fig. 6). The linear discriminant functions (LD1 and LD2) explain more than 90% of the group separation in a 2-D plane. Accuracy of training and testing matrices range from 60% to 100% (Table 4-5, 7-10, 12-15, 17-18 in Supplementary material). WC and EC variables better separate the clusters in winter, spring and fall seasons (Fig. 6 c,d,f, n Table 6-11-16-19 in Supplementary material) when no storm surges are detected. During summer, the contribution of temperature (T) in linear discriminant functions increases (Fig. 6 e,i). This occurs because temperature reaches high values in the low sites in summer months. A general trend can be recognized considering seasons without storm surges. In winter, clusters are separated while in summer they reduce the distances from each-other (Fig. 6b). When storm surges occur, the distance between clusters also decreases (Fig. 6a,b,h,l,m, Fig.7, Fig. S2 in Supplementary material). In these occasions the LDA is not able to clearly separate the groups and the importance of each variable in the linear discriminant functions tends to be equal. In spring and fall of 2020, two storm surges only partially affected the low sites (Fig. 6 d,f). These surges contributed to increase the distance between the low forest clusters and the others, and to increase the ellipses dimension relative to the low forest sites. The effect of the storm surge of September 2020 was still felt in winter 2021, when the cluster distances reached maximum values (Fig. 6g, Fig. 7c). The storm event that occurred in winter 2022 was felt in the low and medium forest. As a consequence, distances between low and medium forest clusters reduced, separating the clusters from the high forest (Fig. 6n). An increase in distance among clusters is regulated by linear discriminant coefficients (Table 6-11-16-19). When coefficients for EC and WC are higher the distance positively increases (winter seasons with no storm surge events), when they decrease the distance decreases too. Mean coefficient trends associated to EC and WC in LD1

and LD2 represents the fluctuations of distances over the seasons. The distance decrease that occurred in summer is correlated to a decrease in the linear discriminant coefficients for both WC and EC (Fig. 7b). During the storm surge event occurred in fall 2020, groundwater conductivity increase is felt in the lower area over the next winter season. Here the coefficient of both WC and EC reached their peaks (Fig. 7b). On the other hand, during storm surge events felt in the entire domain, coefficients are low, because representative of a homogenous condition of shut down where linear discriminant analysis struggles to better separate group. Seasonal trends can be related to cumulative precipitation. During summer, higher rainfall amounts are measured in the study area (Fig. 7a).

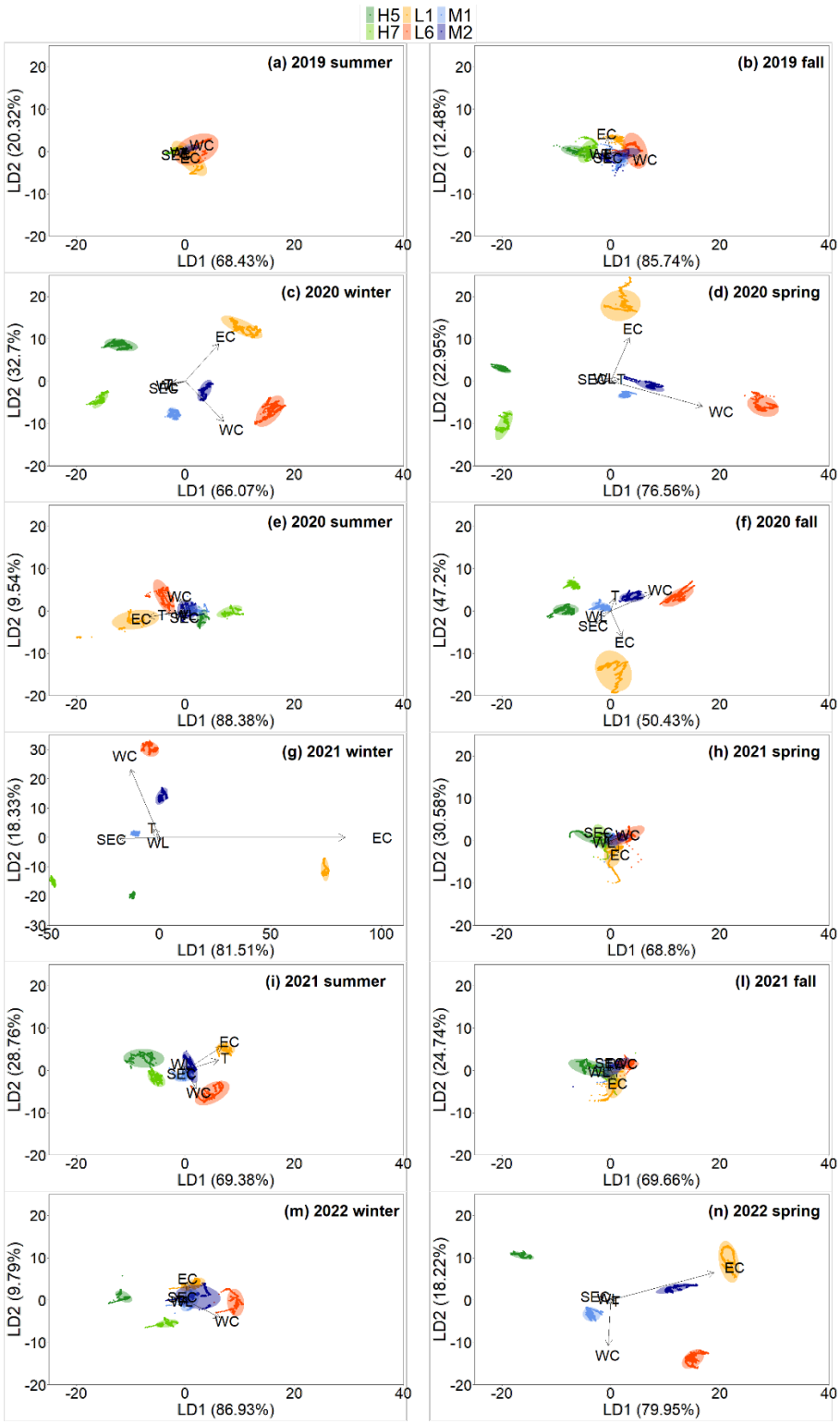

Figure 6: LDA results for each season from 2019 to 2022. Arrows are proportional to the linear function loadings. WL=groundwater level, EC= groundwater electrical conductivity, T=temperature, WC= soil water content, SEC=soil electrical conductivity. Note: in (g) the x- and y-axis are different from the other plots because values are higher.

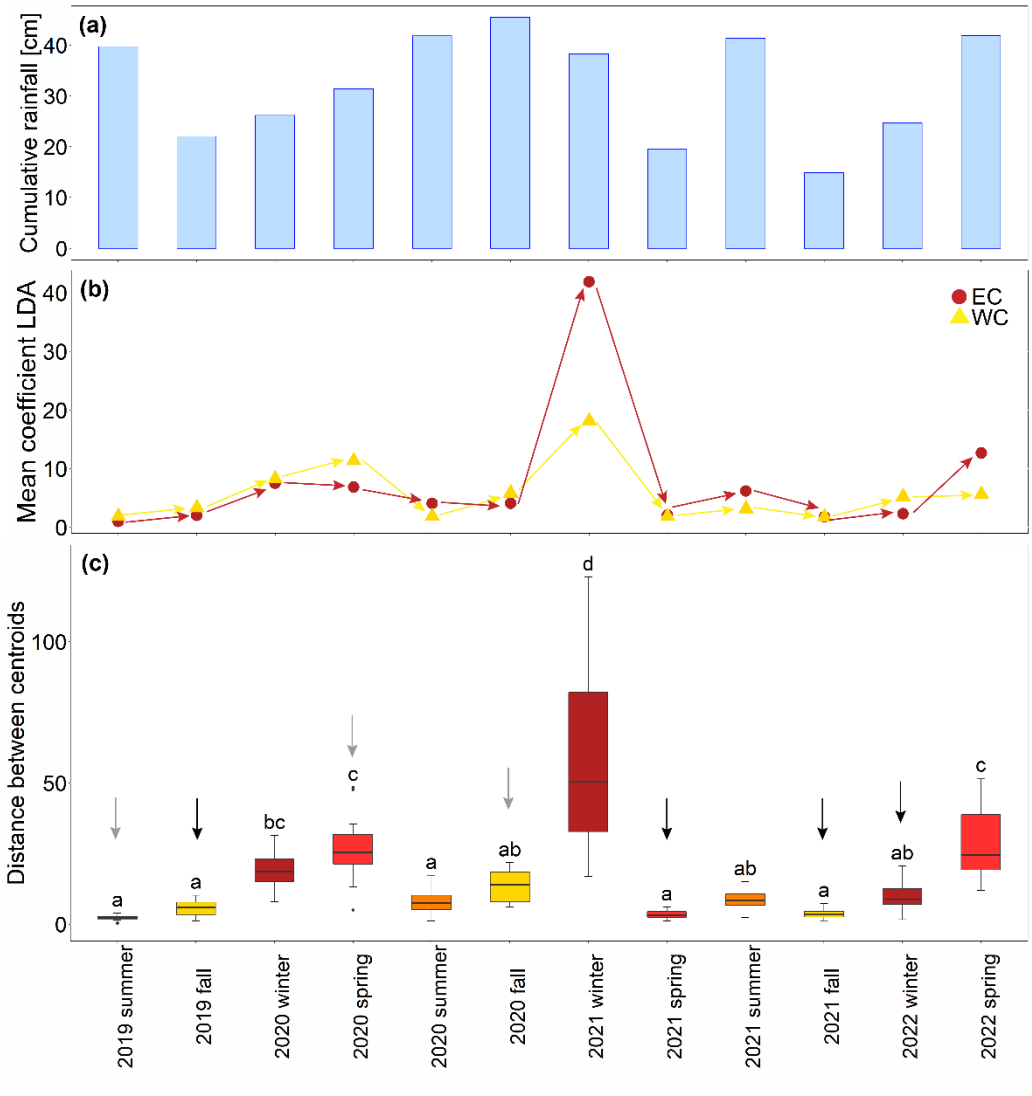

Figure 7: (a) Seasonal cumulative precipitation from 2019 to 2021 collected by NOAA in Norfolk city (VA). (b) Trends of absolute values of the mean coefficient of LD1 and LD2 for WC and EC over the seasons. (c)Euclidean distances between centroids over seasons from 2019 to 2021. Letters represent the post-hoc test results after ANOVA test (90% significance level). Black arrows identify the occurrence of storm surge events in all sites. Gray arrows identify storm surge events only felt in the low sites.

Healthy vs Marsh: classification of points from the groups and storm surge effects

Linear discriminant functions are then calculated using data collected from June 2021 to May 2022 in H5 and Marsh sites. H5 is chosen as representative of 'Healthy forest', since no *Phragmites australis* stems are detected and tree basal area reach its peak here. The "Marsh" site is chosen as

the second end member, where forest vegetation has been fully replaced by marsh vegetation. Both training and testing accuracies are 100% during the period analyzed. Because only two groups are present, one linear discriminant function is used to separate them. The model is used then to classify data from the remaining groups (Fig. 8). H7 station is classified as 'Healthy forest' 87% of the time (Fig. 8a). Storm surge event occurred in fall 2021 (SS1) changes hydrological conditions here, making this site more similar to the Marsh site. This effect is felt for around 1 month. The effects of the storm surge occurred in winter 2022 (SS3) cannot be estimated due to missing data. Similarly, M1 station is classified as 'Healthy forest' 68% of the time (Fig. 8b) and their hydrological conditions are closer to those in the Marsh for around 2 months after the storm surge event in fall 2021 (SS1) and for around 1 month after the storm surge event in winter 2022 (SS3). Station M2 is classified as 'Healthy forest' 52% of the time (Fig. 8c), and storm surges in fall 2021 (SS1) and winter 2022 (SS3) make the hydrological conditions more similar to the Marsh conditions for around 5 months. For both sites, the effect of the storm surge of late spring 2021 (May 30, 2021) are felt. Results for M1 and M2 confirm the intermediate evolution stage of the medium forest in comparison to the high forest. In L1 and L6, the forest is classified as 'Marsh' 99% and 73% of the time respectively (Fig. 8d,e). Here, in addition to SS1 and SS3, another storm surge event is felt in fall 2021 (SS2). In L6 forest is classified as 'Healthy forest' from August to October 2021, when hydrological variables recover from the storm surge event occurred in late spring 2021.

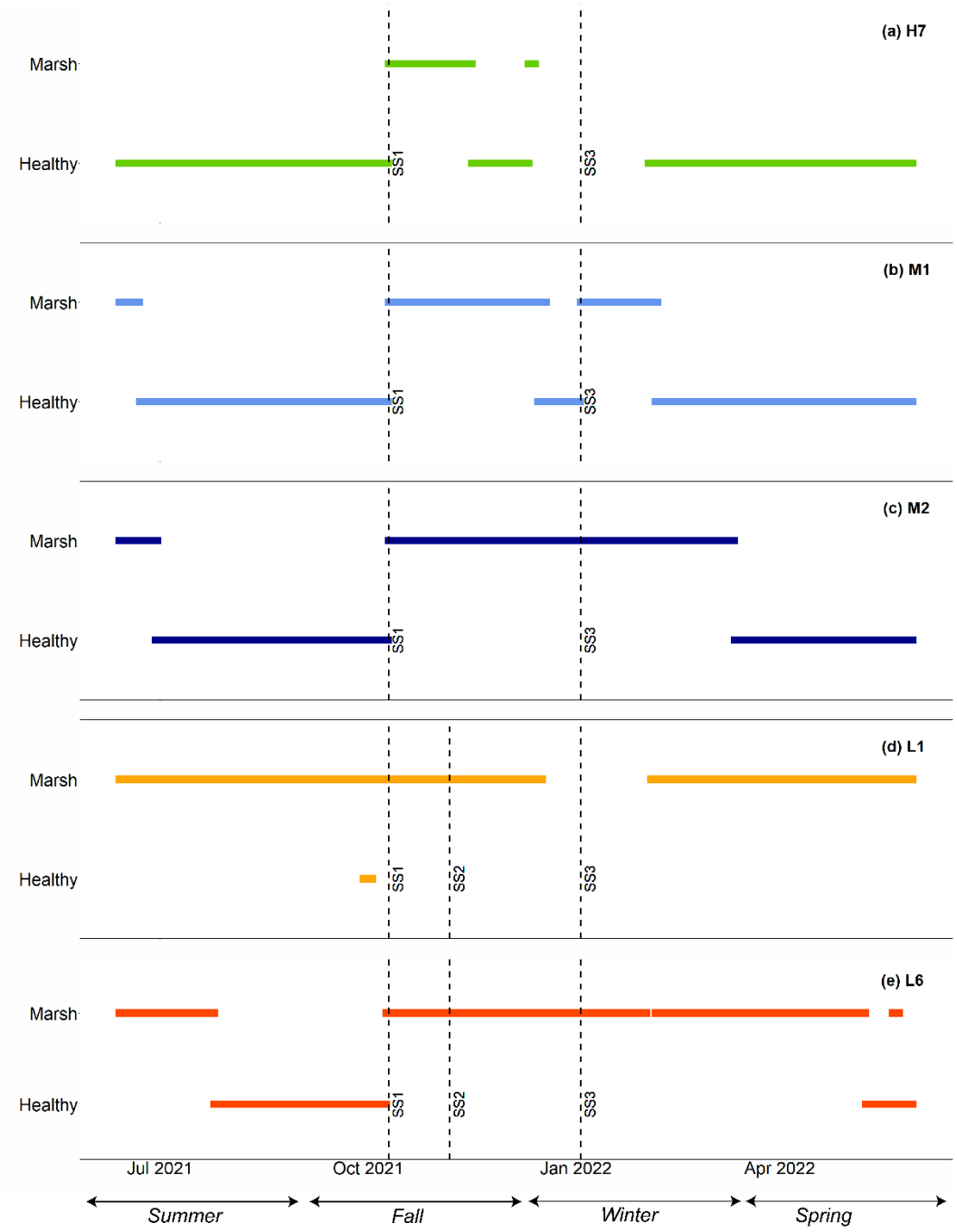

Figure 8: Classification of points from clusters H7, M1, M2, L1 and L6 using Healthy forest-Marsh model derived from linear discriminant analysis. Note: SS1= Storm surge 1, SS2= Storm surge 2, SS3=Storm surge 3.

Discussion

Overall, LDA is a powerful tool to separate and classify clusters based on hydrological variables. This new approach allows to identify the variables most contributing to groups' separation. The analysis of single variables in long-term series is time consuming, while LDA can detect general trends reducing the analysis dimensions to two.

Group separation: feedbacks between hydrology and ecology

Previous works indicate that groundwater levels control plants diversity and forest retreat (Antonellini & Mollema, 2010; Williams et al. 1999; Chen et al. 2006); in our results, the role of WL in the separation of the clusters is negligible. Contrary to forest margins that are sloping (e.g. Nordio and Fagherazzi, 2021) here we could suppose that the flat landscape is characterized by limited hydraulic gradients that do not influence the hydrological budget and consequently ecological zonation. WC and EC are instead the main variables controlling the hydrological diversity of the sites. These variables, expression of flooding events (Unger et al. 2009), are more sensible to the external drivers. The first layer of soil is in fact subject to rapid infiltration and exfiltration when rainfall events and temperature changes occur.

In the absence of storm surges, our results not only confirm a difference in the hydrological variables in the three forested zones, but also between sites in the same zone. This is in accordance with ecological zonation at the forest boundary (Williams et al. 1999a; Williams et al. 1999b). Williams et al. (1999) studied *Sabal palmetto* zonation in a coastal forest in Florida. In their subplots, characterized by a mean salinity of 30 mS/cm and a large number of flooding weeks, the number of dead stands was respectively 3 times and ten times higher than in the intermediate and healthy forest. As a consequence, marsh grass cover increased going from the healthy forest to the

dieback area. Trees basal area is lower in the low forest compared to the high forest, since this parameter is likely correlated to flooding and salinization events. Krauss et al. (2009) showed that Bald cypress (*Taxodium distichum*) basal area was affected by flooding events and salinity values. Frequent flooding and high salinity slows down basal area growth.

In our analysis, lower LD1 values are reached in the high forested area (sites H5 and H7). Here groundwater salinity is on average 5 ppt, the maximum tolerance level where *Pinus taeda* can survive without presenting stress (Pezeshki 1992). Where LD1 is higher, hydrological difference between sites is more evident. Both high EC in L1 and high WC in L6 are toxic for in-situ tree species, likely triggering dieback (Poulter et al. 2008). The reduced forest canopy in the low forest allows the penetration of solar radiation and encourages *Phragmites australis* expansion in late-spring/summer seasons. Shaw et al. (2022), studied *Phragmites australis* encroachment in coastal forests in the Eastern shore of Chesapeake Bay (VA). Although their results suggest that *Phragmites australis* can survive in low-light, this plant prefers high-light conditions. Before marsh vegetation establishment, *Phragmites australis* is the dominant species in areas dominated by saltwater intrusion and flooding events. Therefore, *Phragmites australis* presence represents a first stage in coastal forest retreat.

The effect of external drivers on seasonal patterns

Our results suggest a seasonal trend in hydrological variables. Both rainfall and evapotranspiration seem to improve the groundwater conductivity and soil water content conditions in the low sites during summer, making them closer to high-forest soil conditions. A Higher water content and a reduction in conductivity in the low forest (L1 and L6), could be correlated to rainfall and evapotranspiration during the summer season. Where the water content is higher (L6) the decrease due to evapotranspiration is higher (Ren et al. 2022). Where the conductivity is higher (L1) the

decrease due to rainfall could be higher due to the dilution effect. As consequences, rainfall affects more the medium and low forest, causing a decrease in EC and moving M1, M2, L1 and L6 clusters toward high forest conditions. Moreover, during summer, evapotranspiration increases, mainly due to higher temperature and solar radiation. In low and medium forest sites, the decrease of WC due to evapotranspiration is more significant, reducing the clusters distance from the high forest.

Other reasons could explain this trend in our data. Tree canopies can intercept large amounts of rainfall before they reach the soil (Aston 1979; Xiao & McPherson 2011). Cao et al. (2006) estimated that rainfall interception in a mid-rotation loblolly pine plantation in the coastal plain of North Carolina ranged from 17% to 26%. Being the trees in the low forest barren or dead, more rainfall can reach the soil and infiltrate, creating milder edaphic conditions better suited for trees. In the high forest instead, where trees are healthy, canopy rainfall interception could limit the rainfall amount reaching the soil (Cao et al. 2006). During summer, evapotranspiration is higher mainly due to highr temperatures. This contribute to lower soil water content, which is important for tree roots, that prefer a dry environment (Pezeshki 1991; Poulter et al. 2008). The presence of new invasive vegetation, mostly growing in the summer season in the low forest, could also encourage the water content decrease trough transpiration.

Once this happens, the soil is ready to gather water in the next flooding event. Because plants cannot long live in anoxia conditions, this dynamic is good to encourage soil draining. At the same time, drier soils are more prone to be filled once storm surge events occur, triggering survival of in-situ vegetation. Overall, we can conclude that rainfall and evapotranspiration are essential for the seasonal hydrology of the sites. During summer they can improve the low site hydrological conditions. This is good for both trees and the establishment of new invasive vegetation, exploiting the less harsh environmental conditions.

When storm surge events occur, groundwater level, conductivity, water content, and soil conductivity reach maximum values. If the storm surge affects all sites, the groundwater and soil conditions reach similar values everywhere, justifying cluster collapsing. Storm surge events homogenize hydrological conditions, increasing groundwater conductivity and soil water content in the high forest. The homogenizing process strongly affects the biodiversity and the original ecosystem functioning (Konar et al. 2013). Therefore, these events drive the forest retreat and marsh establishment (Fagherazzi et al. 2019; Kirwan & Gedan, 2019; Kearney et al. 2019). When high conductivity values are reached in the high forest, *Pinus taeda* photosynthetic activity, stomatal conductance, and net carbon assimilation drastically decrease (Munns & Tester, 2008; Pezeshki 1992). When storm surges are only felt in the low forest sites, they make groundwater and soil conditions harsher, differencing these sites from the others.

Hydrological variables in the low forest reach values more similar to those reached in the marsh site. Classification done using linear discriminant analysis with two classes is not only easy to perform, but it clearly shows differences among forested areas and their hydrological closeness to the salt marsh or healthy forest. More frequent storm surge events in the low sites are encouraging the marsh transgression here. Results from the LDA analyses together with the ecological data reinforce this thesis. The medium forest mirrors intermediate conditions of hydrological conditions, and represents an intermediate stage of the transition. Here storm surge events have a stronger impact in terms of EC and WC than in the high forest. Overall, storm surge effects are essential drivers for the forest transgression over the years.

Conclusions

Linear discriminant analysis can easily describe the general trend of hydrological variables along the marsh-forest ecotone. Using this tool, we can: (1) determine the variables that better separate

the hydrology of each site; (2) visualize seasonal hydrological patterns over the years; and (3) identify the effects of storm surge events. Our results show that soil water content and groundwater electrical conductivity are the dominant variables in the linear discriminant functions. Hydrological differences among sites are related to ecological stages of the forest retreat. The feedbacks between hydrology and ecology are essential to understand hydrological conditions in summer. In this season, high evapotranspiration in the low sites combined with significant rainfall decrease groundwater salinity and soil water content, helping the trees to stay alive. Large storm surge events tend to homogenize groundwater and soil conditions in the entire forest boundary, accelerating forest retreat. Small surges affect only the low forest, enhancing the hydrological difference between the low forest sites and the others. Hydro-ecological zonation is confirmed by the classification analysis using LDA, where low forest is classified more than 80% of the time as salt marsh, indicating that salt marsh transgression is occurring in these areas, driven by more frequent storm surges.

References

1. Allen, J. A., Conner, W. H., Goyer, R. A., Chambers, J. L., & Krauss, K. W. (1998). Freshwater forested wetlands and global climate change. *Vulnerability of coastal wetlands in the Southeastern United States, A Biological Science Report USGS/BRD/BSR, Hawaii*, 33-44.
2. Antonellini, M., & Mollema, P. N. (2010). Impact of groundwater salinity on vegetation species richness in the coastal pine forests and wetlands of Ravenna, Italy. *Ecological Engineering*, *36*(9), 1201-1211.
3. Aston, A. R. (1979). Rainfall interception by eight small trees. *Journal of hydrology*, *42*(3-4), 383-396.
4. Balakrishnama, S., & Ganapathiraju, A. (1998). Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, *18*(1998), 1-8.
5. Cai, J., Liu, Y., Lei, T., & Pereira, L. S. (2007). Estimating reference evapotranspiration with the FAO Penman–Monteith equation using daily weather forecast messages. *Agricultural and Forest Meteorology*, *145*(1-2), 22-35.
6. Cao, W., Sun, G., McNulty, S. G., Chen, J., Noormets, A., Skaggs, R. W., & Amatya, D. M. (2006). Evapotranspiration of a mid-rotation loblolly pine plantation and a recently harvested stands on the coastal plain of North Carolina, USA. In *Hydrology and Management of Forested Wetlands, Proceedings of the International Conference, April 8-12, 2006, New Bern, North Carolina* (p. 5). American Society of Agricultural and Biological Engineers.
7. Chambers, R. M., Meyerson, L. A., & Saltonstall, K. (1999). Expansion of Phragmites australis into tidal wetlands of North America. *Aquatic botany*, *64*(3-4), 261-273.
8. Chen, Y. N., Zilliacus, H., Li, W. H., Zhang, H. F., & Chen, Y. P. (2006). Ground-water level affects plant species diversity along the lower reaches of the Tarim river, Western China. *Journal of Arid Environments*, *66*(2), 231-246.
9. Church, J. A., & White, N. J. (2006). A 20th century acceleration in global sea-level rise. *Geophysical research letters*, *33*(1).
10. Correa-Metrio, A., Cabrera, K. R., & Bush, M. B. (2010). Quantifying ecological change through discriminant analysis: a paleoecological example from the Peruvian Amazon. *Journal of Vegetation Science*, *21*(4), 695-704.
11. Dasgupta, S., Laplante, B., Murray, S., & Wheeler, D. (2009). Sea-level rise and storm surges: A comparative analysis of impacts in developing countries. *World Bank Policy Research Working Paper*, (4901).
12. Desantis, L. R., Bhotika, S., Williams, K., & Putz, F. E. (2007). Sea-level rise and drought interactions accelerate forest decline on the Gulf Coast of Florida, USA. *Global Change Biology*, *13*(11), 2349-2360.
13. Doyle, T. W., Krauss, K. W., Conner, W. H., & From, A. S. (2010). Predicting the retreat and migration of tidal forests along the northern Gulf of Mexico under sea-level rise. *Forest Ecology and Management*, *259*(4), 770-777.
14. Fagherazzi, S. and G. Nordio. 2022. Groundwater, soil moisture, light and weather data in Brownsville forest, Nassawadox, VA, 2019-2022. Virginia Coast Reserve Long-Term Ecological Research Project Data Publication knb-lter-vcr.349.4 (doi:10.6073/pasta/942a5a981e6e986c5fa1a9a9cd2eb8b7).Gardner, L. R., Reeves, H. W., & Thibodeau, P. M. (2002). Groundwater dynamics along forest-marsh transects in a southeastern salt marsh, USA: Description, interpretation and challenges for numerical modeling. *Wetlands Ecology and Management*, *10*(2), 143-157.
15. Gedan, K., J. MacGregor, R. Mohammadi and A. Khan. 2021. Forest Transition Experiment - Vegetation Monitoring on a Coastal Virginia Forest, 2019-2021. Virginia Coast Reserve Long-

Term Ecological Research Project Data Publication knb-lter-vcr.361.3 (doi:10.6073/pasta/f5d6779021676538649c445dc3a4fbf7).
16. Geng, X., Heiss, J. W., Michael, H. A., Boufadel, M. C., & Lee, K. (2020). Groundwater flow and moisture dynamics in the swash zone: Effects of heterogeneous hydraulic conductivity and capillarity. *Water Resources Research*, *56*(11), e2020WR028401.
17. Gleeson, T., Marklund, L., Smith, L., & Manning, A. H. (2011). Classifying the water table at regional to continental scales. *Geophysical Research Letters*, *38*(5).
18. Gotelli, N. J., & Ellison, A. M. (2004). *A primer of ecological statistics* (Vol. 1). Sunderland: Sinauer Associates.
19. Izenman, A. J. (2013). Linear discriminant analysis. In *Modern multivariate statistical techniques* (pp. 237-280). Springer, New York, NY.
20. Kearney, W. S., Fernandes, A., & Fagherazzi, S. (2019). Sea-level rise and storm surges structure coastal forests into persistence and regeneration niches. *PloS one*, *14*(5), e0215977.
21. Kirwan, M. L., & Gedan, K. B. (2019). Sea-level driven land conversion and the formation of ghost forests. *Nature Climate Change*, *9*(6), 450-457.
22. Konar, M., Todd, M. J., Muneepeerakul, R., Rinaldo, A., & Rodriguez-Iturbe, I. (2013). Hydrology as a driver of biodiversity: Controls on carrying capacity, niche formation, and dispersal. *Advances in Water Resources*, *51*, 317-325.
23. Krauss, K. W., Duberstein, J. A., Doyle, T. W., Conner, W. H., Day, R. H., Inabinette, L. W., & Whitbeck, J. L. (2009). Site condition, structure, and growth of baldcypress along tidal/non-tidal salinity gradients. *Wetlands*, *29*(2), 505-519.
24. Michael, H. A., Russoniello, C. J., & Byron, L. A. (2013). Global assessment of vulnerability to sea-level rise in topography-limited and recharge-limited coastal groundwater systems. *Water Resources Research*, *49*(4), 2228-2240.
25. Mohamed, A. A., Sasaki, T., & Watanabe, K. (2000). Solute transport through unsaturated soil due to evaporation. *Journal of Environmental Engineering*, *126*(9), 842-848.
26. Monteith, J., & Unsworth, M. (2013). *Principles of environmental physics: plants, animals, and the atmosphere*. Academic Press.
27. Mueller-Dombois, D., & Ellenberg, H. (1974). Vegetation types: a consideration of available methods and their suitability for various purposes.
28. Munns, R., & Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.*, *59*, 651-681.
29. Miyamoto, S. (2008). *Salt tolerance of landscape plants common to the southwest*. Texas Water Resources Institute.
30. Nachshon, U. (2018). Cropland soil salinization and associated hydrology: Trends, processes and examples. *Water*, *10*(8), 1030.
31. NOAA National Centers for Environmental information, Climate at a Glance: County Time Series, published May 2022, retrieved on May 11, 2022 from https://www.ncdc.noaa.gov/cag/
32. Nordio, G., & Fagherazzi, S. (2022). Salinity increases with water table elevation at the boundary between salt marsh and forest. *Journal of Hydrology*, *608*, 127576.
33. Paldor, A., & Michael, H. A. (2021). Storm surges cause simultaneous salinization and freshening of coastal aquifers, exacerbated by climate change. *Water Resources Research*, *57*(5), e2020WR029213.
34. Pennings, S. C., & Callaway, R. M. (1992). Salt marsh plant zonation: the relative importance of competition and physical factors. *Ecology*, *73*(2), 681-690.
35. Pezeshki, S. R. (1992). Response of Pinus taeda L to soil flooding and salinity. In *Annales des sciences forestières* (Vol. 49, No. 2, pp. 149-159). EDP Sciences.
36. Poulter, B., Christensen, N. L., & Qian, S. S. (2008). Tolerance of Pinus taeda and Pinus serotina to low salinity and flooding: Implications for equilibrium vegetation dynamics. *Journal of Vegetation Science*, *19*(1), 15-22.

37. Ren, X., Zhang, Q., Yue, P., Yan, X., & Yang, Y. (2022). Energy Distribution Characteristics and Environmental Impact Factors in Different Regions of the Summer Monsoon Transition Zone. *Frontiers in Environmental Science*, 344.
38. Rencher, A. C., & Christensen, W. F. (2012). Methods of Multivariate Analysis.--John Wiley & Sons, Inc.
39. Rice, J., Ohmart, R. D., & Anderson, B. W. (1983). Habitat selection attributes of an avian community: a discriminant analysis investigation. *Ecological Monographs*, *53*(3), 263-290.
40. Sallenger, A. H., Doran, K. S., & Howd, P. A. (2012). Hotspot of accelerated sea-level rise on the Atlantic coast of North America. *Nature Climate Change*, *2*(12), 884-888.
41. Schieder, N. W., & Kirwan, M. L. (2019). Sea-level driven acceleration in coastal forest retreat. *Geology*, *47*(12), 1151-1155.
42. Shaw, P., Jobe, J., & Gedan, K. B. (2022). Environmental limits on the spread of invasive Phragmites australis into upland forests with marine transgression. *Estuaries and Coasts*, *45*(2), 539-550.
43. Smith, J. A. (2013). The role of Phragmites australis in mediating inland salt marsh migration in a mid-Atlantic estuary. *PloS one*, *8*(5), e65091.
44. Stewart, S., Ivy, M. A., & Anslyn, E. V. (2014). The use of principal component analysis and discriminant analysis in differential sensing routines. *Chemical Society Reviews*, *43*(1), 70-84.
45. Sumner, D. M., & Belaineh, G. (2005). Evaporation, precipitation, and associated salinity changes at a humid, subtropical estuary. *Estuaries*, *28*(6), 844-855.
46. Tebaldi, C., Strauss, B. H., & Zervas, C. E. (2012). Modelling sea level rise impacts on storm surges along US coasts. *Environmental Research Letters*, *7*(1), 014032.
47. Tharwat, A. (2016). Linear vs. quadratic discriminant analysis classifier: a tutorial. *International Journal of Applied Pattern Recognition*, *3*(2), 145-180.
48. Thibodeau, P. M., Gardner, L. R., & Reeves, H. W. (1998). The role of groundwater flow in controlling the spatial distribution of soil salinity and rooted macrophytes in a southeastern salt marsh, USA. *Mangroves and Salt Marshes*, *2*(1), 1-13.
49. Tully, K., Gedan, K., Epanchin-Niell, R., Strong, A., Bernhardt, E. S., BenDor, T., ... & Weston, N. B. (2019). The invisible flood: The chemistry, ecology, and social implications of coastal saltwater intrusion. *BioScience*, *69*(5), 368-378.
50. Wiegleb, G. (1981). Application of multiple discriminant analysis on the analysis of the correlation between macrophyte vegetation and water quality in running waters of Central Europe. *Hydrobiologia*, *79*(1), 91-100.
51. Williams, K., Ewel, K. C., Stumpf, R. P., Putz, F. E., & Workman, T. W. (1999). Sea-level rise and coastal forest retreat on the west coast of Florida, USA. *Ecology*, *80*(6), 2045-2063.
52. Williams, K., Pinzon, Z. S., Stumpf, R. P., & Raabe, E. A. (1999). Sea-level rise and coastal forests on the Gulf of Mexico. *US Geological Survey*, *1500*, 20910.
53. Wu, W., Mallet, Y., Walczak, B., Penninckx, W., Massart, D. L., Heuerding, S., & Erni, F. (1996). Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Analytica Chimica Acta*, *329*(3), 257-265.
54. Xanthopoulos, P., Pardalos, P. M., & Trafalis, T. B. (2013). Linear discriminant analysis. In *Robust data mining* (pp. 27-33). Springer, New York, NY.
55. Xiao, Q., & McPherson, E. G. (2011). Rainfall interception of three trees in Oakland, California. *Urban Ecosystems*, *14*(4), 755-769.
56. Yang, J., Zhang, H., Yu, X., Graf, T., & Michael, H. A. (2018). Impact of hydrogeological factors on groundwater salinization due to ocean-surge inundation. *Advances in Water Resources*, *111*, 423-434.
57. Yu, X., Yang, J., Graf, T., Koneshloo, M., O'Neal, M. A., & Michael, H. A. (2016). Impact of topography on groundwater salinization due to ocean surge inundation. *Water Resources Research*, *52*(8), 5794–5812. https://doi.org/10.1002/2016wr018814

Supplementary material

Table 1: Accuracy matrices for training and testing data in 2020.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 96%	**H5**	5045	15	0	0	13	73
	H7	181	5288	0	0	0	0
	L1	0	0	5244	0	0	0
	L6	0	0	35	4834	1	0
	M1	59	0	0	0	4959	39
	M2	0	0	76	432	301	5221

Accuracy testing 96%	**H5**	3353	9	0	0	7	44
	H7	100	3472	0	0	0	0
	L1	0	0	3368	0	0	0
	L6	0	0	25	3227	0	0
	M1	45	0	0	0	3288	30
	M2	0	0	36	291	214	3376

Table 2: Accuracy matrices for training and testing data in 2021.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 96%	**H5**	4516	328	0	0	0	12
	H7	753	3388	0	122	795	660
	L1	0	557	4777	24	302	313
	L6	0	1	0	4981	0	281
	M1	0	247	0	1	3834	447

	M2	0	526	351	159	322	3588

Accuracy testing 96%	**H5**	2976	224	0	0	1	8
	H7	515	2185	0	99	547	414
	L1	0	385	3040	22	194	208
	L6	0	0	0	3222	0	216
	M1	0	156	0	4	2558	301
	M2	0	347	242	126	207	2312

Table 3: coefficients of linear classifiers for each variable in 2020 and 2021.

	2020		**2021**	
Original Variables	**LD1**	**LD2**	**LD1**	**LD2**
WL	-0.79	-1.04	-1.23	0.10
EC	2.25	-2.87	-0.10	2.75
T	0.54	-0.49	-0.60	-0.25
WC	3.54	2.26	2.83	-0.69
SEC	-1.24	0.17	-0.79	-1.61

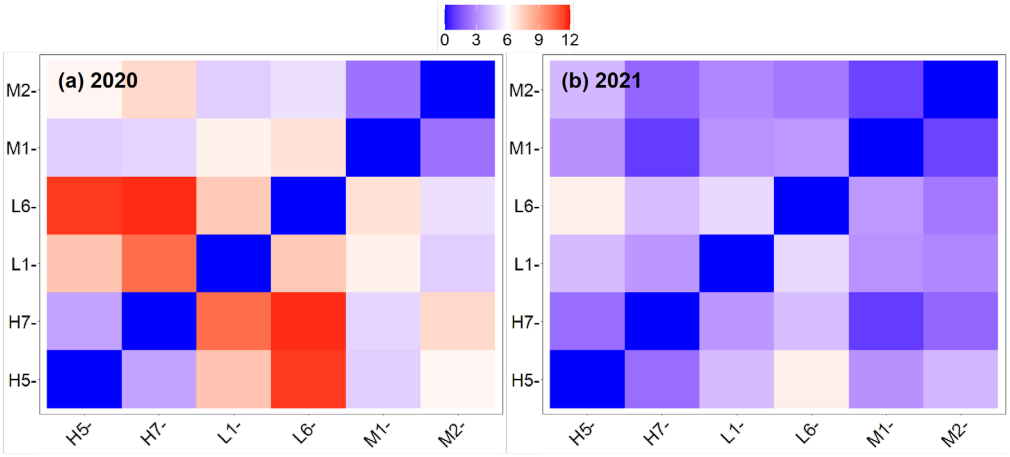

Figure S1: Euclidean distances between centroids in 2020 (a) and 2021 (b)

.

Table 4: Accuracy matrices for training and testing data in summer 2019.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 63%	**H5**	1138	254	7	96	0	32
	H7	234	959	196	35	2	2
	L1	0	0	455	232	0	0
	L6	0	0	67	607	2	176
	M1	0	120	603	38	902	67
	M2	0	0	28	341	433	1046

Accuracy testing 65%	**H5**	701	166	10	68	0	20
	H7	159	637	128	29	2	0
	L1	0	0	311	135	0	0
	L6	0	0	43	431	0	115
	M1	0	96	362	24	624	37
	M2	0	0	22	196	267	737

Table 5: Accuracy matrices for training and testing data in fall 2019.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 84%	**H5**	1056	238	0	0	0	0
	H7	264	1022	1	0	5	93
	L1	0	0	1270	5	3	0
	L6	0	0	0	1059	0	64
	M1	0	0	0	212	1026	62
	M2	0	0	37	3	262	1035

Accuracy testing 83%	**H5**	651	209	0	0	0	0
	H7	165	667	1	0	3	56
	L1	0	0	811	4	5	0
	L6	0	0	0	705	0	47
	M1	0	0	0	148	677	46
	M2	0	0	16	0	155	733

Table 6: coefficients of linear classifiers for each variable for each season in 2019.

	Summer		**Fall**	
Original Variables	**LD1**	**LD2**	**LD1**	**LD2**
WL	-0.61	-0.12	-1.54	-0.39
EC	0.76	-1.23	-0.60	3.45
T	-0.18	-0.25	-0.65	-0.47
WC	2.53	1.26	5.22	-1.52
SEC	-1.44	-0.62	-0.41	-1.17

Table 7: Accuracy matrices for training and testing data in winter 2020.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 100%	**H5**	1345	0	0	0	0	0
	H7	0	1293	0	0	0	0
	L1	0	0	1331	0	0	0
	L6	0	0	0	1309	0	0
	M1	0	0	0	0	1316	0
	M2	0	0	0	0	0	1298

Accuracy testing 100%	**H5**	838	0	0	0	0	0
	H7	0	891	0	0	0	0
	L1	0	0	853	0	0	0
	L6	0	0	0	875	0	0
	M1	0	0	0	0	868	0
	M2	0	0	0	0	0	886

Table 8: Accuracy matrices for training and testing data in spring 2020.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 100%	**H5**	1372	0	0	0	0	0
	H7	0	1333	0	0	0	0
	L1	0	0	1356	0	0	0
	L6	0	0	0	1349	0	0
	M1	0	0	0	0	1339	0
	M2	0	0	0	0	0	1323

Accuracy testing 100%	**H5**	860	0	0	0	0	0
	H7	0	899	0	0	0	0
	L1	0	0	876	0	0	0
	L6	0	0	0	883	0	0
	M1	0	0	0	0	893	0
	M2	0	0	0	0	0	909

Table 9: Accuracy matrices for training and testing data in summer 2020.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 92%	**H5**	1371	0	0	0	108	40
	H7	0	1333	0	0	0	0
	L1	0	0	1356	0	0	0
	L6	0	0	0	1349	0	0
	M1	1	0	0	0	904	150
	M2	0	0	0	0	327	1133

Accuracy testing 92%	**H5**	860	0	0	0	71	30
	H7	0	899	0	0	0	0
	L1	0	0	876	0	0	0
	L6	0	0	0	882	0	0
	M1	0	0	0	0	607	96
	M2	0	0	0	1	215	783

Table 10: Accuracy matrices for training and testing data in fall 2020.

	Actual

	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 100%	**H5**	1372	0	0	0	0	0
	H7	0	1333	0	0	0	0
	L1	0	0	1356	0	0	0
	L6	0	0	0	1349	0	0
	M1	0	0	0	0	1339	0
	M2	0	0	0	0	0	1323

Accuracy testing 100%	**H5**	860	0	0	0	0	0
	H7	0	899	0	0	0	0
	L1	0	0	876	0	0	0
	L6	0	0	0	883	0	0
	M1	0	0	0	0	893	0
	M2	0	0	0	0	0	909

Table 11: coefficients of linear classifiers for each variable for each season in 2020.

	Winter		**Spring**		**Summer**		**Fall**	
Original Variables	**LD1**	**LD2**	**LD1**	**LD2**	**LD1**	**LD2**	**LD1**	**LD2**
WL	-2.79	-0.75	-0.96	0.57	-0.03	-0.79	-2.23	-0.99
EC	6.15	8.86	3.38	10.32	-6.72	-1.47	2.12	-6.07
T	-2.58	-0.53	1.59	0.49	-3.54	-0.69	0.70	3.04
WC	7.02	-9.51	16.84	-5.94	-0.86	2.94	7.54	4.16
SEC	-3.30	-1.33	-2.81	0.42	-0.12	-1.26	-2.72	-2.90

Table 12: Accuracy matrices for training and testing data in winter 2021.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 100%	**H5**	1333	0	0	0	0	0
	H7	0	1278	0	0	0	0
	L1	0	0	1328	0	0	0
	L6	0	0	0	1290	0	0
	M1	0	0	0	0	1300	0
	M2	0	0	0	0	0	1272

Accuracy testing 100%	**H5**	827	0	0	0	0	0
	H7	0	882	0	0	0	0
	L1	0	0	832	0	0	0
	L6	0	0	0	870	0	0
	M1	0	0	0	0	860	0
	M2	0	0	0	0	0	888

Table 13: Accuracy matrices for training and testing data in spring 2021.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 79%	**H5**	1086	10	0	0	0	4
	H7	286	1012	0	0	18	6
	L1	0	108	1321	1	64	0
	L6	0	0	0	1109	0	246
	M1	0	203	0	171	1175	361
	M2	0	0	35	68	82	706
Accuracy testing 80%	**H5**	697	10	0	0	0	1
	H7	163	684	0	0	15	2
	L1	0	81	864	3	43	0
	L6	0	0	0	713	0	172
	M1	0	124	0	127	782	236
	M2	0	0	12	40	53	498

Table 14: Accuracy matrices for training and testing data in summer 2021.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 96%	**H5**	1372	0	0	0	0	0
	H7	0	1333	0	0	29	0
	L1	0	0	1356	0	0	0
	L6	0	0	0	1346	0	0
	M1	0	0	0	3	1138	119
	M2	0	0	0	0	172	1204
Accuracy testing 96%	**H5**	860	2	0	0	0	0
	H7	0	897	0	0	6	0
	L1	0	0	876	0	0	0
	L6	0	0	0	883	0	0
	M1	0	0	0	0	748	73
	M2	0	0	0	0	139	836

Table 15: Accuracy matrices for training and testing data in fall 2021.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 79%	**H5**	1143	64	0	0	40	21
	H7	71	929	79	0	216	226
	L1	0	23	1008	2	77	0
	L6	0	1	0	1137	0	4
	M1	105	155	0	1	791	21
	M2	1	4	161	141	162	994
Accu	**H5**	699	70	0	0	29	18

	H7	36	598	58	1	136	156
	L1	0	11	646	0	37	0
	L6	0	0	0	766	0	1
	M1	80	128	0	0	554	4
	M2	1	1	98	88	94	691

Table 16: coefficients of linear classifiers for each variable for each season in 2021.

	Winter		**Spring**		**Summer**		**Fall**	
Original Variables	**LD1**	**LD2**	**LD1**	**LD2**	**LD1**	**LD2**	**LD1**	**LD2**
WL	-0.57	-1.45	-1.05	-0.29	-0.54	1.31	-1.81	-0.18
EC	83.71	0.04	1.41	-2.77	6.73	5.64	1.05	-2.54
T	-2.94	2.77	-0.84	0.67	6.09	2.48	-0.34	1.18
WC	-13.11	23.23	2.57	1.18	2.01	-4.33	2.20	1.38
SEC	-18.12	-0.50	-1.85	1.64	-0.56	-0.66	-0.17	1.52

Table 17: Accuracy matrices for training and testing data in winter 2022.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 93%	**H5**	1268	0	0	0	0	0
	H7	0	668	0	0	0	0
	L1	0	0	643	0	0	0
	L6	0	0	0	1254	0	0
	M1	0	0	6	0	880	15
	M2	0	0	0	6	396	1234
Accuracy testing 93%	**H5**	791	0	0	0	0	0
	H7	0	447	0	0	0	0
	L1	0	0	440	0	0	0
	L6	0	0	0	796	0	0
	M1	0	0	1	0	546	7
	M2	0	0	0	4	289	878

Table 18: Accuracy matrices for training and testing data in spring 2022.

		Actual					
	Predicted	**H5**	**H7**	**L1**	**L6**	**M1**	**M2**
Accuracy training 100%	**H5**	1054	0	0	0	0	0
	H7	0	1022	0	0	0	0
	L1	0	0	997	0	0	0
	L6	0	0	0	1041	0	0
	M1	0	0	0	0	1034	0
	M2	0	0	0	0	0	992

Accuracy testing 100%	**H5**	662	0	0	0	0	0
	H7	0	672	0	0	0	0
	L1	0	0	695	0	0	0
	L6	0	0	0	651	0	0
	M1	0	0	0	0	658	0
	M2	0	0	0	0	0	699

Table 19: coefficients of linear classifiers for each variable for each season in 2022.

	Winter		**Spring**	
Original Variables	**LD1**	**LD2**	**LD1**	**LD2**
WL	-0.53	-0.18	-0.42	0.30
EC	0.36	4.27	18.90	6.48
T	-0.85	0.00	0.78	-0.19
WC	6.37	-4.13	-0.47	-10.81
SEC	-0.46	0.61	-2.36	0.71

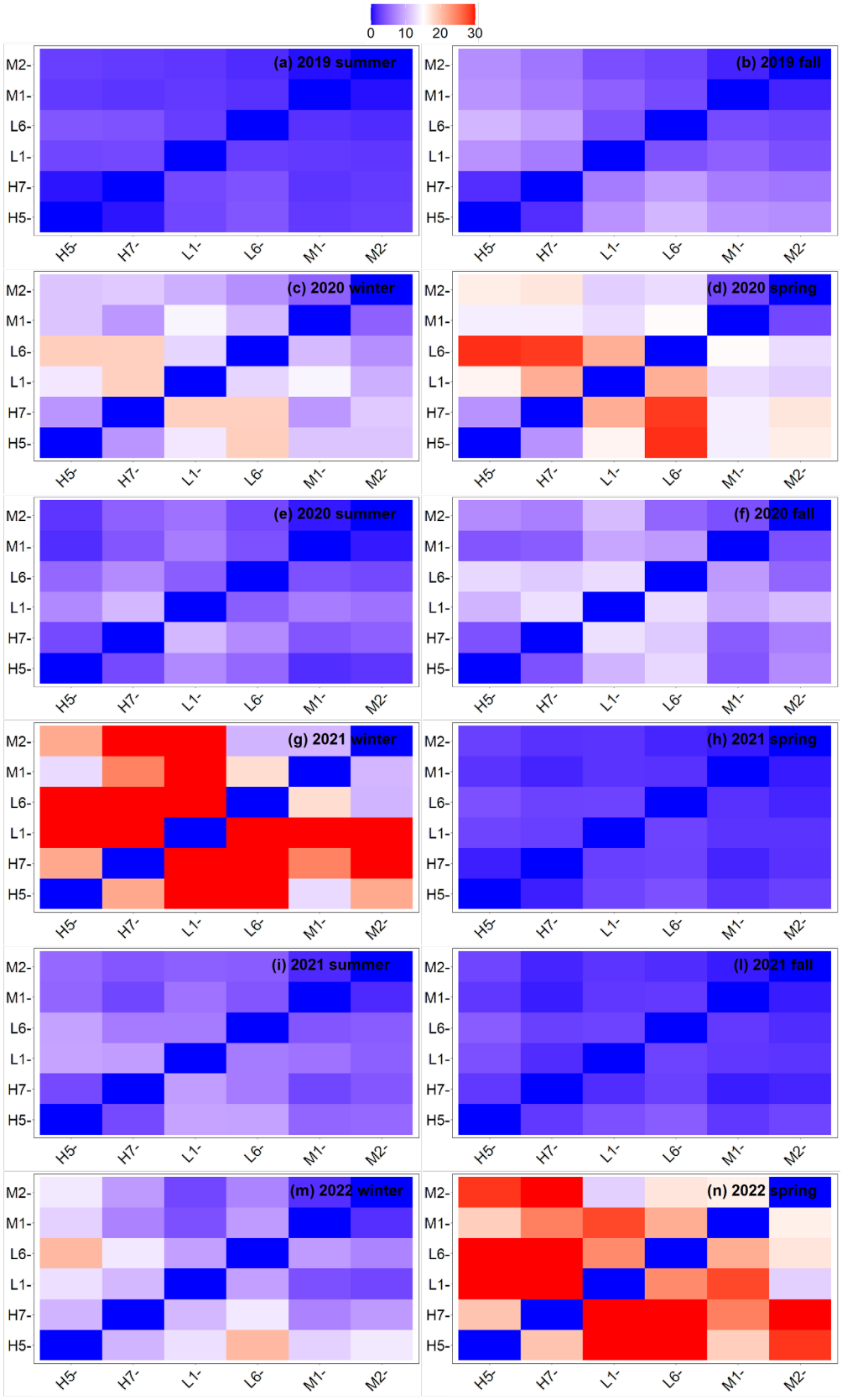

Figure S2: Euclidean distances between centroids over seasons from 2019 to 2022.