**מעבדה במידול מערכות אקולוגיות – תרגיל בית 2**

|  |  |  |
| --- | --- | --- |
| **משימות שהושלמו** | **משימות שהוקצו** | **שם חבר הצוות** |
| כל המשימות הושלמו. | הגדרת מערכת אקולוגיות, ניסוח השערות והגדרת משתנים, סימולציה והדמיה. | אושר בן זזון |
| כל המשימות הושלמו. | ניתוח רב משתני באמצעות PCA, פיתוח מודל סטטיסטי מרחבי, סימולציה והדמיה. | אנדי לואיס ספנר (מהנדס מערכת) |

קישור לתיקיית ה-Git: <https://github.com/Andy-Lewis-Sapner/EML_EcoCoders.git>

חלק א': הגדרת מערכת אקולוגית

(1) נחל הקישון (בתחום רשות נחל הקישון) הכולל את גדותיו, מימיו, בעלי החיים והצומח שנמצאים בסביבה זו.

(2)

* רכיבים ביוטיים (מינים עיקריים): סך כל האוכלוסיות במערכת (קיים חוסר נתונים עבור כמויות של כל אוכלוסייה בצורה מדויקת למשך זמן).
* רכיבים אביוטיים (גורמים פיזיקליים וכימיים): טמפרטורת מים, חומציות (pH), ריכוז חמצן מומס, עכירות (opacity), מוליכות חשמלית.
* מקורות אנרגיה: רכיבים ביוטיים (בעיקר בקשרי טורף/נטרף), שמש (קרינה), מי הנחל (זרימה).
* מעגלי חומרים (לפחות שניים):
  + פחמן (C) (צמחים קולטים CO₂ בפוטוסינתזה, צרכנים משחררים CO₂ בנשימה, מפרקים מחזירים CO₂ לאטמוספירה).
  + חנקן (N) (מגיע מהקרקע או מזיהום, משמש לבניית חלבונים ביצורים חיים, חיידקים ממחזרים חנקן בתהליך דה-ניטריפיקציה).
* קשרים ומשובים בין הרכיבים:
  + קשר גומלין טורף-נטרף: (יותר טורפים = פחות טרף, פחות טורפים = יותר טרף) - קשר הופכי.
  + חמצן ופעילות ביולוגית: (יותר פוטוסינתזה ← יותר חמצן במים ← יותר דגים) - קשר ישיר.
  + עכירות ופעילות ביולוגית: (פחות פוטוסינתזה = פחות אור = עכירות) - קשר הופכי.

חלק ב': ניסוח השערות והגדרת משתנים

(1) השערה 1: האם יש קשר ישיר או הופכי בין כל אחד מהמדדים הקיימים והנבחרים במערכת לבין סך כל כמות הפרטים בכל האוכלוסיות?  
השערה 2: האם יש תחום כלשהו של סך כל כמות הפרטים בכל האוכלוסיות שניתן לצפות שלא נחצה (גם כלפי מעלה וגם כלפי מטה)?

(2)

* משתנים תלויים: סך כל כמות הפרטים בכל האוכלוסיות, חיידקים קוליפורמים.
* משתנים בלתי תלויים: ערך הגבה pH, חמצן מומס, טמפרטורה, עכירות, מוליכות חשמל ספציפית.
* משתנים מתערבים שיש ביכולתכם לבקר: חום גבוה, רוח חזקה.

(3)

נשער כי הקשרים הקיימים בין המשתנים הוא כלהלן:

קשר ישיר, קשר הופכי, קשר אפס.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| מוליכות חשמל ספציפית | עכירות | טמפרטורה | חמצן מומס | pH | חיידקים קוליפורמים | סה"כ פרטים | **סוג קשרים בין המשתנים** |
| הופכי | הופכי | הופכי | ישיר | אפס | הופכי | - | סה"כ פרטים |
| ישיר | ישיר | ישיר | הופכי | אפס | - | הופכי | חיידקים קוליפורמים |
| ישיר | אפס | אפס | ישיר | - | אפס | אפס | pH |
| הופכי | הופכי | הופכי | - | ישיר | הופכי | ישיר | חמצן מומס |
| ישיר | ישיר | - | הופכי | אפס | ישיר | הופכי | טמפרטורה |
| ישיר | - | ישיר | הופכי | אפס | ישיר | הופכי | עכירות |
| - | ישיר | ישיר | הופכי | ישיר | ישיר | הופכי | מוליכות חשמל ספציפית |

חלק ג': ניתוח רב-משתני באמצעות PCA

(1)

* מטרת ה-PCA היא להוריד את כמות המשתנים שייעשה בהם שימוש. זאת בכדי להשאיר את הדפוסים והקשרים החשובים ביותר ביניהם. מכיוון שהמחקר הנוכחי כולל 6 משתנים סביבתיים (ומשתנה המייצג תצפיות), נוכל להשתמש בPCA כדי להשאיר את המשתנים המשפיעים.
* PCA יכיל את המשתנים הבאים: רווית חמצן מומס, עכירות, ערך הגבה, מוליכות חשמלית ספציפית, טמפרטורה, חיידקים קוליפורמים וכמות תצפיות ידועות.
* נייצר גרפים מתאימים ל-PCA, ונשתמש בPC1 ו-PC2 (שהם בעלי ערכי השונות הגבוהים ביותר), כדי לבדוק מהם המשתנים המשפיעים ביותר.

(2)

* גרף הbiplot ישתמש ב-PC1 ו-PC2 ויציג וקטורים בעלי אורכים וזוויות של כל הנתונים ויראה איזה משתנה (או משתנים) משפיעים ביותר על נקודות זמן בנתונים שלנו.
* נבדוק אילו משתנים מתקבצים ליד אילו וקטורים, וכך נזהה אילו משתנים הם המשפיעים ביותר בנתונים. היחסים ניתנים לזיהוי באמצעות בדיקת הזוויות השונות בין כל שני ווקטורים המייצגים נתונים.

חלק ד': פיתוח מודל סטטיסטי מרחבי

1. מודל סטטיסטי שניתן להשתמש בו הוא רגרסיה, מכיוון שקיימים הרבה משתנים, ולפי ניתוח ה-PCA, ניתן לראות כי רוב הנקודות מתקבצות ליד המשתנים – חיידקים קוליפורמים, עכירות, מוליכות חשמלית ספציפית ואוכלוסיות.
2. נשתמש במודל Kriging כמודל מרחבי, שיאפשר לבדוק את משתני התצפיות והחיידקים הקוליפורמים במרחב נחל הקישון. כדי לעשות זאת, נשתמש במיקומי התצפיות וגם מיקומי תחנות המדידה בנחל הקישון.
3. ביטוי הדינמיקה המרחבית של המערכת ניתן על ידי בדיקה של רמות התצפיות ומדידות החיידקים בסביבת נחל הקישון. מכיוון שקיים חוסר בנתונים אמיתיים במדידות של נחל הקישון וגם בתצפיות, למרות זיהוי רכיבים עיקריים ב-PCA (שניתן לזהות על ידי התקבצויות בגרף ה-biplot), נעשה שימוש בכל המשתנים. המודל יאפשר לזהות תופעות אקולוגיות כמו מדידות או כמות תצפיות במיקומים שונים, אשר עבורם לא קיים מידע מהימן ומספיק.

חלק ה': סימולציה והדמיה

(1) להלן קישור למחברת Google Colab: [HW2\_EcoCoders](https://colab.research.google.com/drive/1W5kulnfLG_uiSlOH_2HhydUCOYKTwrwf?usp=sharing).

המחברת מכילה 3 כרטיסיות:

כרטיסיה 1 (DataFrames): כרטיסיה זו מציגה את 4 טבלאות הנתונים בהם נעשה שימוש בDashboard שנוצר – מידע עבור PCA, מידע עבור Kriging (שתי טבלאות – אחת לתצפיות ואחת למדידות בנחל) ומידע עבור משתנים מתערבים. בנוסף, ישנם שני גרפים המציגים את המידע של המשתנים המתערבים במשך תקופת 2022 – 2024.

כרטיסיה 2 (PCA Plots): כרטיסיה זו מציגה את הגרפים של PCA. היא מכילה Dropdown עם כל האפשרויות לגרפים השונים, ובבחירה של הגרף המתאים, ייטען הגרף שנבחר.

כרטיסיה 3 (Kriging Plots): כרטיסיה זו מציגה את הגרפים של Kriging. ישנם 4 גרפים, שניים לגבי תצפיות ושניים לגבי מדידות. ישנו Dropdown המאפשר לבחור אחד מתוך שלושת התרחישים (תרחיש בסיסי, תרחיש בלחץ ותרחיש בשיקום). כמו כן, Dropdown נוסף מאפשר לבחור על איזו מדידה להתמקד בכל תרחיש.

(כדי להריץ ולצפות ב-Dashboard, יש להריץ את כל המחברת, ולפתוח התא האחרון תחת הכותרת App וניתן ללחוץ על הקישור שמופיע כדי לראות את ה-Dashboard בכרטיסיה נפרדת בדפדפן).

(2) ב-Dashboard שנוצר במחברת ה-Google Colab מופיעים 3 התרחישים. ניתן לגשת אליהם באמצעות הרצת המחברת ופתיחת הכרטיסייה Kriging Plots.

(3)

* השערה 1 – קשר בין מדדים לבין אוכלוסיות – מצאנו כי קיים קשר, לפעמים הופכי ולפעמים ישיר, בתלות במדד.

השערה 2 – תחום לחציית כמות אוכלוסייה – ראינו כי לפי תצפיות מחודשים שונים, קיימות אוכלוסיות אשר עבורן לא היו תצפיות כלל וכאלו שהיו תצפיות רבות.

* ניתן לראות כי עכירות ומוליכות חשמלית משפיעים אחרת על כל אחד מסוגי האוכלוסיות שאותן ניתן לבדוק – על אוכלוסיות של בעלי חיים וצמחים, ככל שעכירות גבוהה יותר, האוכלוסייה יורדת, אך ככל שהמוליכות גבוהה יותר, גם האוכלוסייה גבוהה יותר. לעומת זאת, קיים קשר שלילי בין חיידקים קוליפורמים לבין עכירות ומוליכות חשמלית – ככל שהם יורדים, החיידקים קוליפורמים. ניתן לראות גם קשר שלילי בין חיידקים קוליפורמים לבין אוכלוסיות (באמצעות מטריצת הקורלציות).
* באמצעות מודל Kriging, כאשר יש מספיק נתונים בסביבה (למשל עבור מדדים של נחל הקישון), נוכל לחזות מרחבית את המדדים הללו במקומות אחרים במים. במקרה וחסרים נתונים (כמו נתוני אוכלוסיות), קשה יהיה לחזות רק באמצעות מודל Kriging מרחבית.