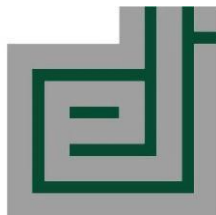


Konsolidētās kovariācijas un to pielietojumi biotopu analīzē

Romāns Dinuls, Aivars Lorencs,
Ints Mednieks

ELEKTRONIKAS UN
DATORZINĀTŅU
INSTITŪTS



INSTITUTE OF
ELECTRONICS AND
COMPUTER SCIENCE

Motivācija

- Liela nozīme statistikā un dzīvē ir jēdzienam «**atkarība**».
- Pētījumos statistiskie modeļi lielākoties tiek būvēti, izmantojot pieņēmumu par **neatkarību**.
 - Piem.: gadījuma iztvērums (jeb gadījuma izlase) ir **neatkarīgu** gadījumu lielumu kopa.
- Toties pieņēmums par **neatkarību** ne vienmēr apraksta realitāti pietiekoši labi.
- Šādos gadījumos bieži vien svarīgi izpētīt, kāda tieši ir **atkarība** starp lielumiem.
- "Kā" vai "caur ko" pētāmās parādības ir saistītas vai mijiedarbojas savā starpā?

Ievads

- **Kas ir kovariācija?** No latīņu: *covariant*.
 - «**Co**» ir prefikss ar nozīmi «**kopīgi, savstarpēji, vienkopus**».
 - «**Variant**» no darbības vārda «*variare*», kurš nozīmē «**mainīties**».
- «**Kovariācija**» burtiskā tulkojumā ir kopējas vai savstarpējas izmaiņas.
- Statistikā par kovariāciju sauc izmaiņu rādītāju, kuru definē sekojoši:

Ievads

- Kovariācijas definīcija:

Kovariācija starp diviem kopīgi sadalītiem reāliem gadījumlielumiem X un Y ar galīgiem 2-ās kārtas momentiem ir skaitlis

$$\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$$

– Piezīme: Ja kovariācija ir 0, tas nenozīmē, ka nav saistības.

- Matemātiskais jēdziens «kovariācija» apzīmē sagaidāmo vidējo vērtību reizinājumam, kurš sastāv no doto gadījumlielumu novirzēm no to vidējām vērtībām.

Ievads

- **Novērtējumam** no iztvēruma datiem kopīgi sadalītiem gadījumlīelumiem X un Y , parasti izmanto šādu lielumu:

$$\widehat{Cov}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}, \text{ kur}$$

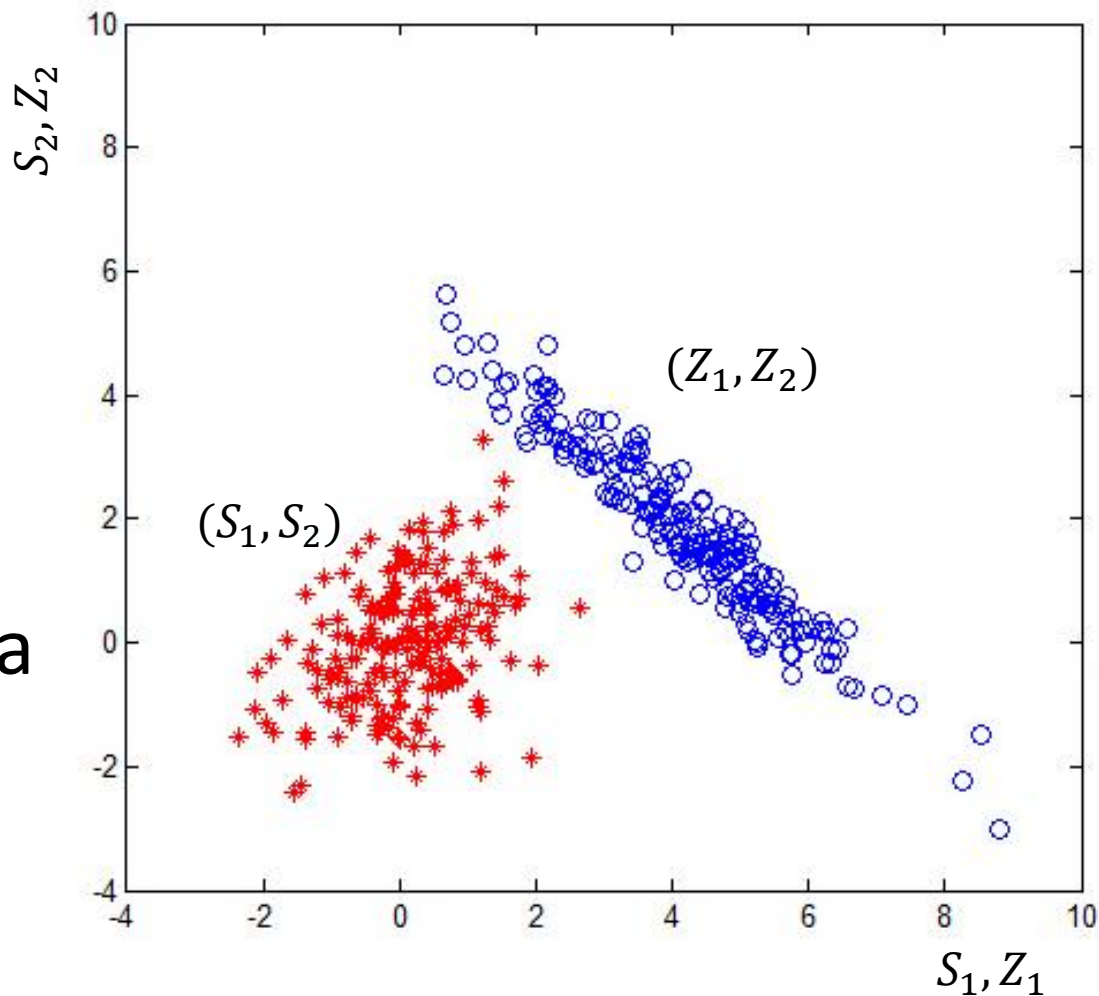
$\bar{x}$ un $\bar{y}$ ir attiecīgie aritmētiskie vidējie,

Vērtības x_i un y_i ir gadījumlīelumu X un Y vienas kopīgas realizācijas $\begin{pmatrix} x_i \\ y_i \end{pmatrix}$ komponentes.

n ir iztvēruma apjoms.

Ievads

- Piemērs:
Dažādi kopējo
izmaiņu varianti:
 (S_1, S_2) un
 (Z_1, Z_2)
statistiskā atkarība
ir citādāka.



Ievads

- **Kovariāciju matrica** gadījumielumu vektoram

$$(X_1, X_2, \dots, X_N)$$

(jāizpildās arī iepriekš minētiem nosacījumiem priekš Cov)
ir matrica:

$$\Sigma = \begin{pmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_N) \\ Cov(X_2, X_1) & \cdots & & \\ \vdots & \cdots & & \vdots \\ Cov(X_N, X_1) & \cdots & & Cov(X_N, X_N) \end{pmatrix}$$

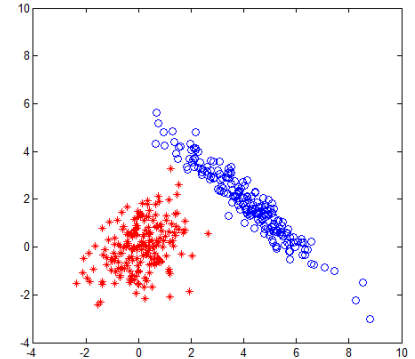
T. i., matricas Σ elementi s_{ij} ir uzdoti sekojoši:

$$s_{ij} = Cov(X_i, X_j)$$

Piezīme: $Cov(X, X) = \mathbb{D}X$ (ir X dispersija).

Ievads

- Īpašības: kovariāciju matrica Σ ir
 - Simetriska
 - Pozitīvi definita (ja nav singulāra)
- Kovariāciju matrica ir summārais rādītājs attiecīgo mainīgo kopējo izmaiņu tendencēm (vai mainīgo savstarpējai saistībai).
- Kovariācijas jēdziena vispārinājums. Satur kovariāciju visiem iespējamām gadījumu lielumu vektora komponentu pāriem.



Ievads

- Piemērs

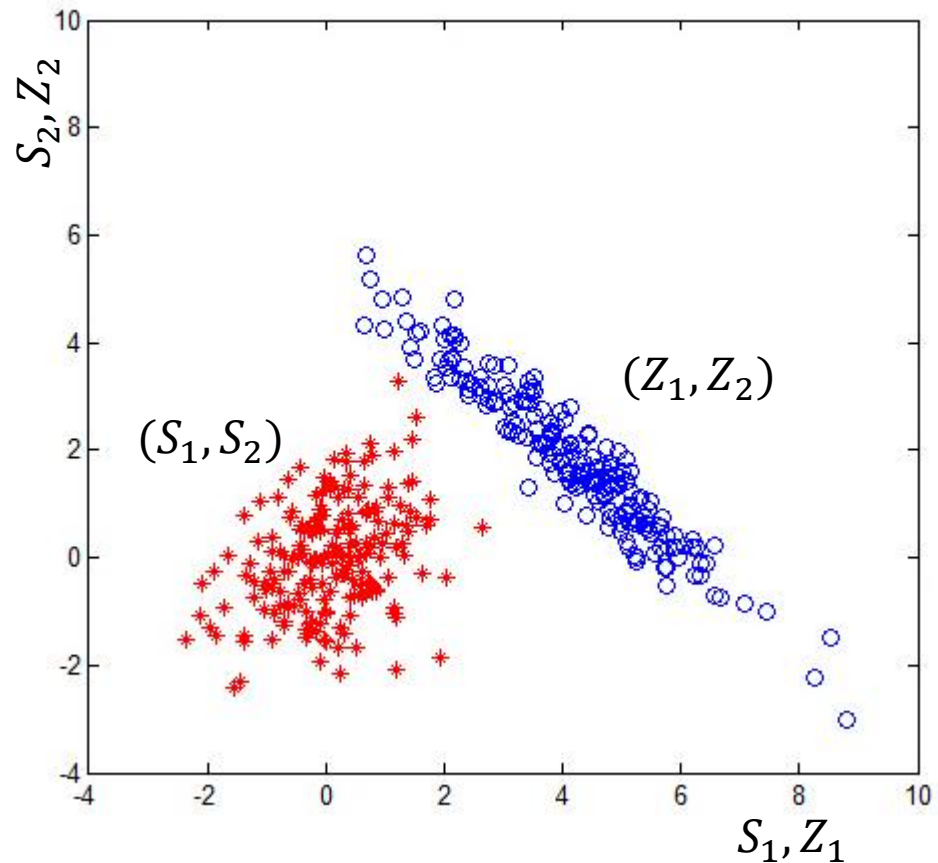
(S_1, S_2) sarkanas «*»:

$$\begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$$

Annotations: $\mathbb{D}S_1$ points to the top-left element (1); $\text{Cov}(S_1, S_2)$ points to the top-right element (0.3); $\text{Cov}(S_2, S_1)$ points to the bottom-left element (0.3); $\mathbb{D}S_2$ points to the bottom-right element (1).

(Z_1, Z_2) zilie «o»:

$$\begin{pmatrix} 2 & -1.9 \\ -1.9 & 2 \end{pmatrix}$$



Ievads

- **Kovariāciju matricas novērtējumam** no iztvēruma datiem parasti izmanto formulu:

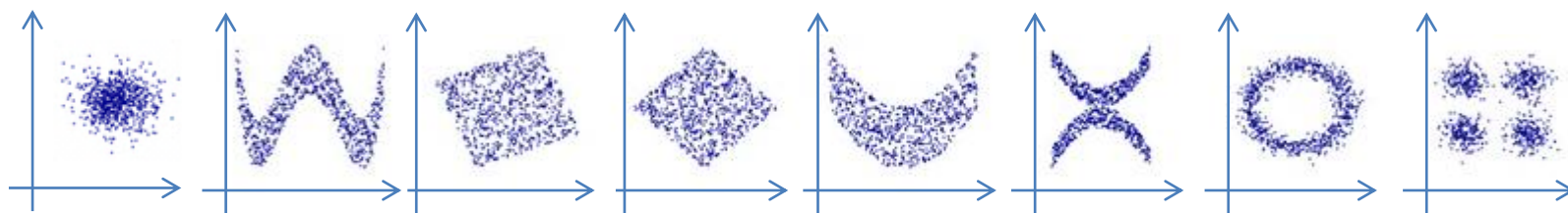
$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T,$$

Kur $\mathbf{x}_i$ ir iztvēruma elements reprezentēts kā kolonnas

vektors $\mathbf{x}_i = \begin{pmatrix} x_{1i} \\ \vdots \\ x_{Ni} \end{pmatrix}$. Savukārt $\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_N \end{pmatrix}$ ir iztvēruma vidējo vektors.

Ievads

- Kovariāciju matrica **neņem** vērā vairākus faktoros. Piemēram, sekojošām atkarībām kovariāciju matricas ir savstarpēji vienādas:



- Lai gan «kovariācija» tulkojams kā «savstarpējas vai kopējas izmaiņas», reāli matemātiskais jēdziens «kovariācija» apraksta tikai daļu no savstarpējām vai kopējām izmaiņām.

Dabas parks «Dvietes paliene».

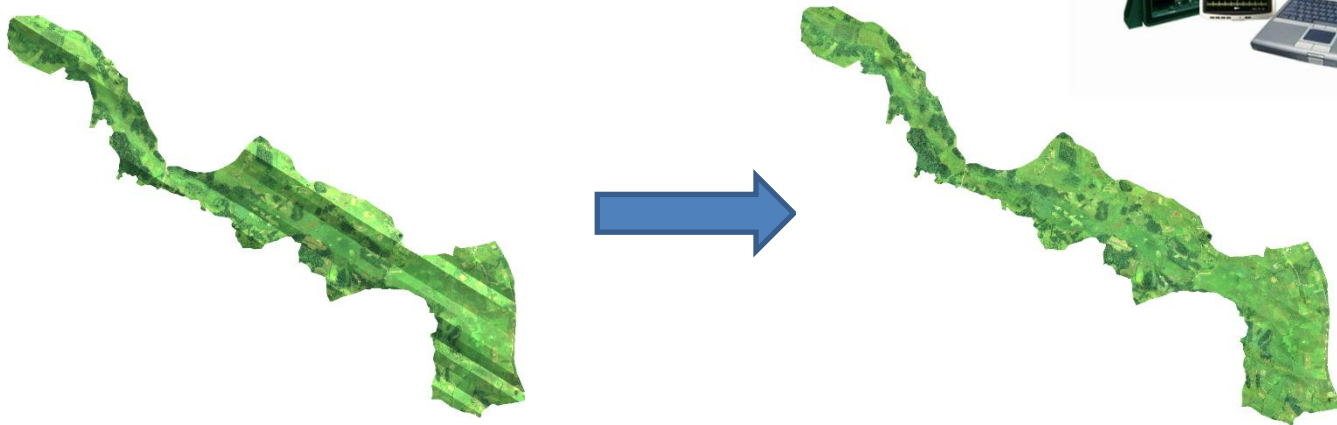
Uzdevums.

Dabas parks «Dvietes paliene»

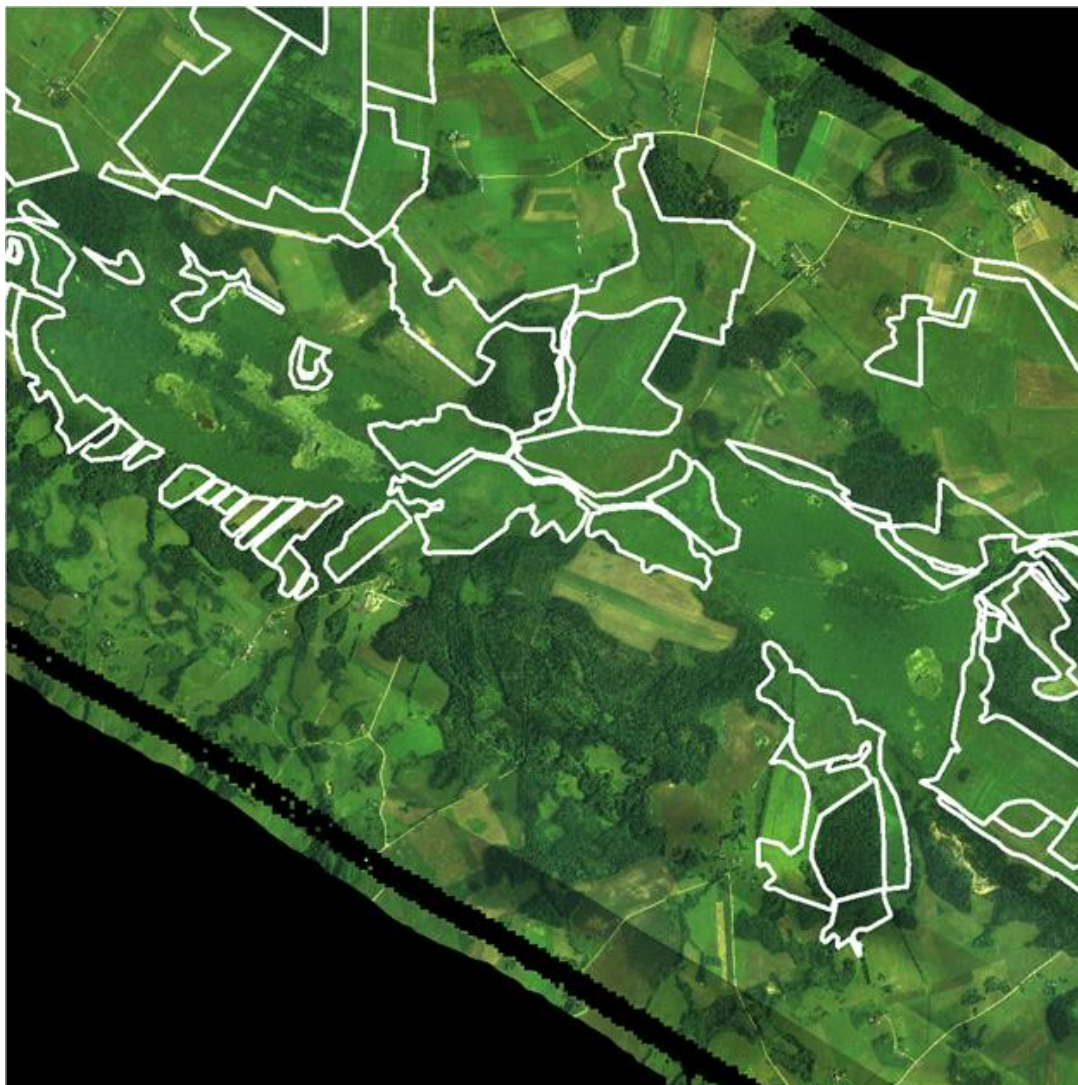


Datu ieguve

- **Multispektrālais** attēls iegūts, izmantojot lidmašīnu.
- Hiperspektrālais sensors.
- Dati savākti pa lidojumu līnijām, pēc tam izlīdzināti.



Attēls tuvumā



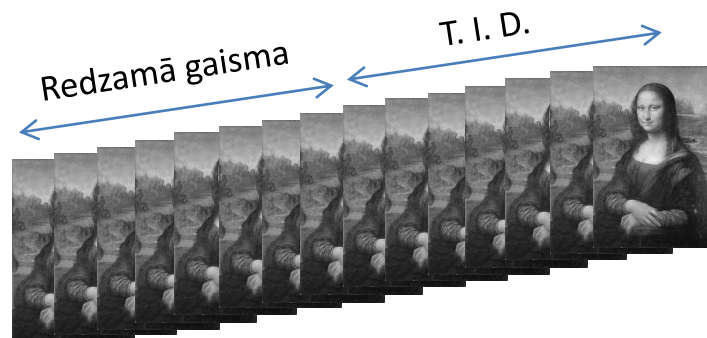
Datu struktūra

- **Multispektrālais attēls** – sastāv no vairākiem slāņiem.
- Katrs pikselis uz attēla ir reprezentēts ar 15 vērtībām (nevis 3, kā parastajai fotogrāfijai).
- Katra slāņa vērtība atbilst spektrālajai joslai. Spektrālās joslas tiek paņemtas no redzamās gaismas un tuvā infrasarkanā diapazona.

Piemērs:

Dažādi slāņi
multispektrālajam
attēlam.

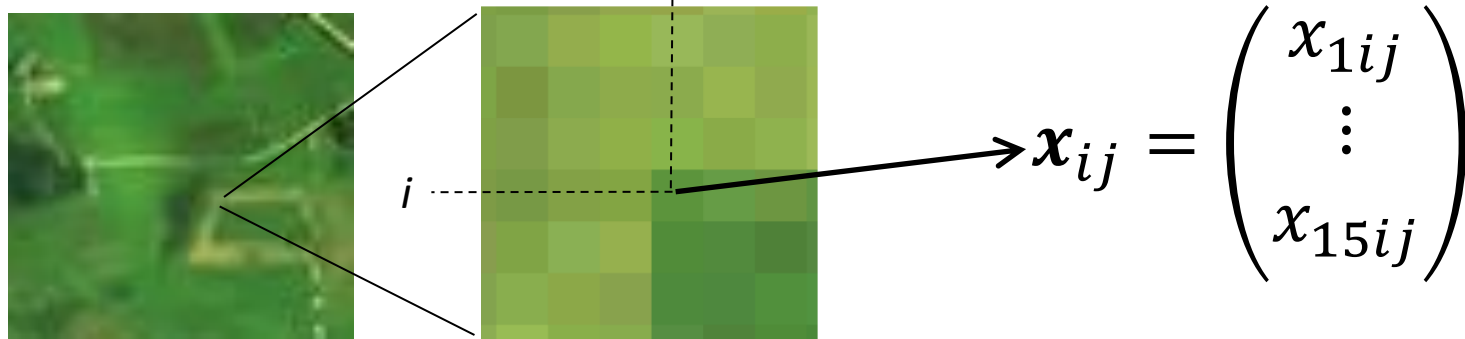
Katrs slānis nes
pildusinformāciju.



Datu matemātiskais modelis

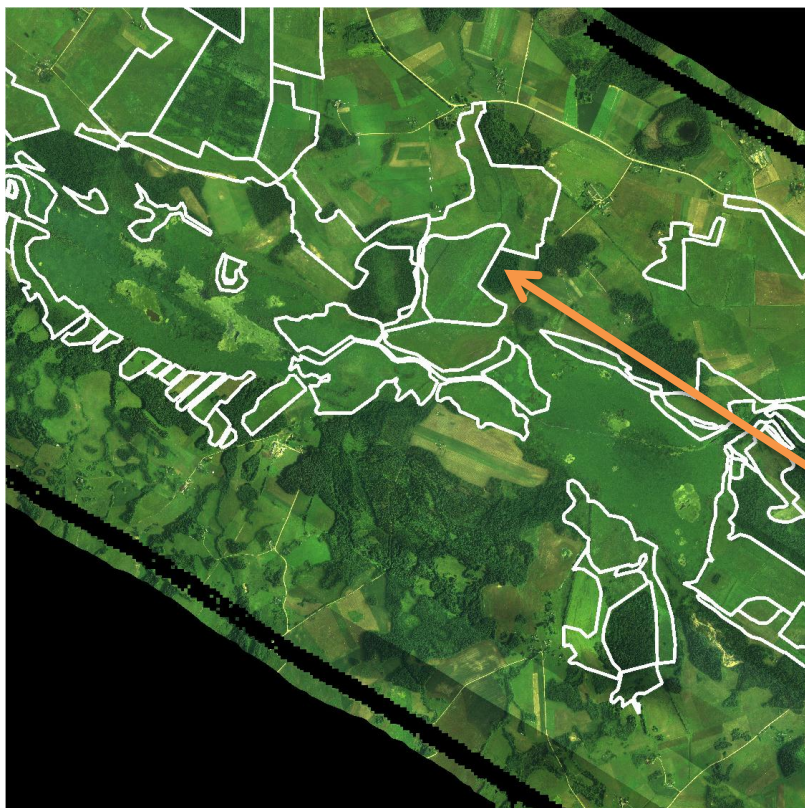
- Pikseļiem uz multispektrālā attēla ir piešķirti indeksi: rindas numurs i un kolonas numurs j .
- Pikselis ar indeksu ij ir reprezentēts ar $N=15$ komponentu kolonu vektoru

$$\mathbf{x}_{ij} = \begin{pmatrix} x_{1ij} \\ \vdots \\ x_{15ij} \end{pmatrix}$$



Biotopti dabas parkā „Dvietes paliene”

- Galvenā interesējošo biotopu atpazīšanas pazīme: interesējošie biotopi ir nekultivētas pļavas, kurās ir liela augu sugu daudzveidība.

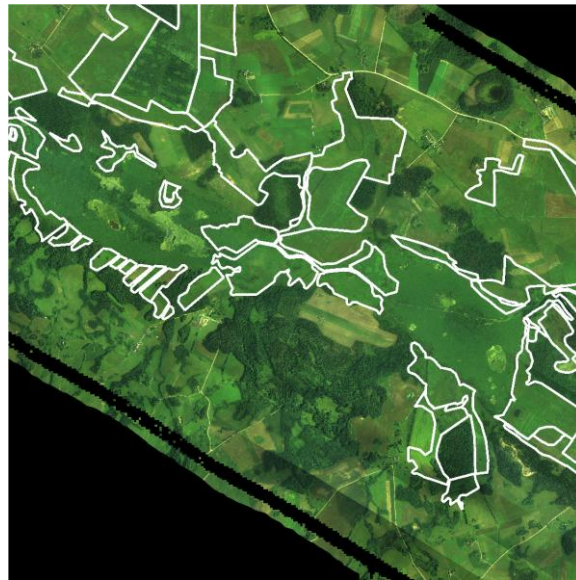


2005. gadā iegūtās
biotopu robežas ir
izdalītas ar baltām
līnijām



Uzdevums

- Pēdējie dati par biotopu robežām iegūti 2005. gadā.



- Uzdevums:
 - **lezīmēt interesējošus biotopus uz attēla, kas iegūts 2011. gadā. Atrast biotopus, kuros notikušas izmaiņas.**

Metode

- Interesējošus biotopus atšķir sekojošas galvenās pazīmes:
 - Interesējošie biotopi ir pļavas
 - Interesējošos biotopos nav dominējošu augu sugu



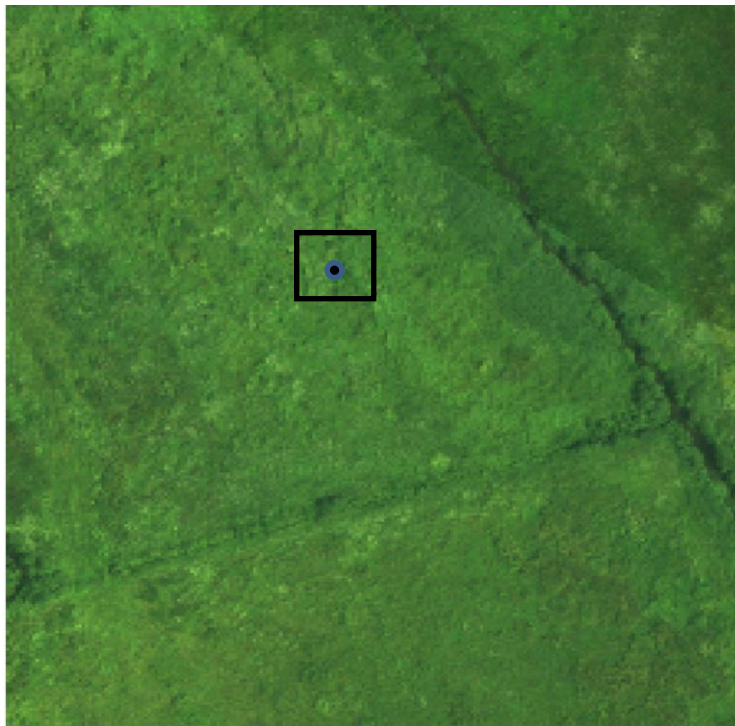
Metode

- **Ideja:**
 - 1) Atrast, kurās vietās uz attēla ir liela augu sugu daudzveidība
 - 2) Atdalīt visus objektus, kuri nav pļavas
- Liela augu sugu **daudzveidība** datos izpaužas sekojošā veidā:
 - Biotopu reģionos ir liela **pikseļu «krāsu» dažādība**, salīdzinot ar citām pļavām

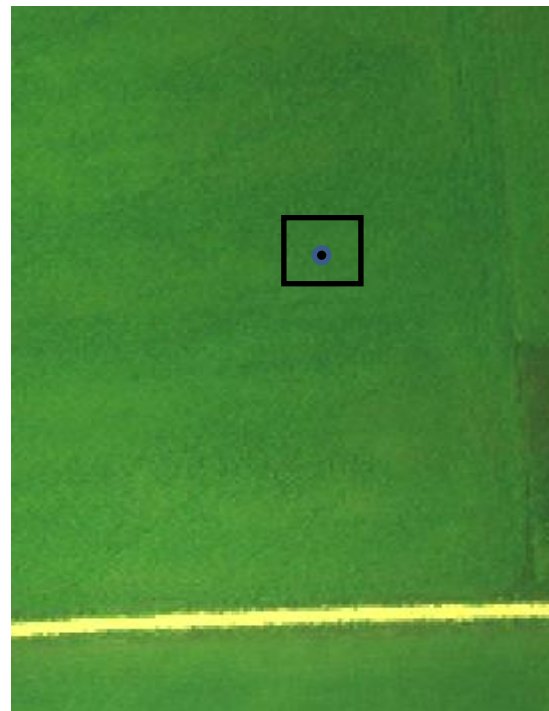
Metode

- Atšķirība starp pļavu, kura ir biotops un citu pļavu. Ir redzamas atšķirības krāsu daudzveidībā, ja tiek apskatīts neliels apgabals ap katru punktu.

Interesējošais biotops



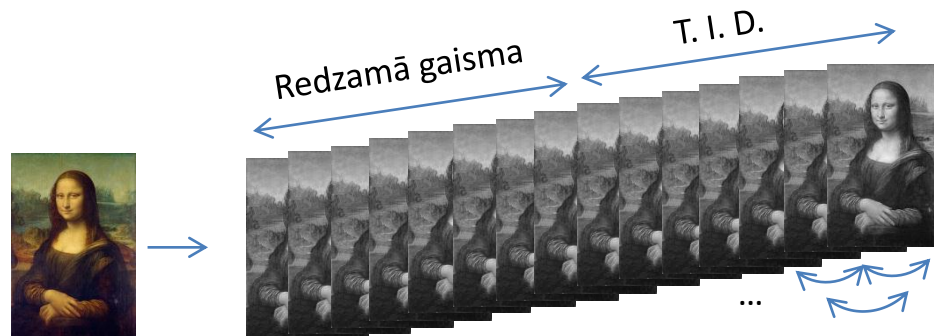
Cita pļava



Konsolidētā kovariācija

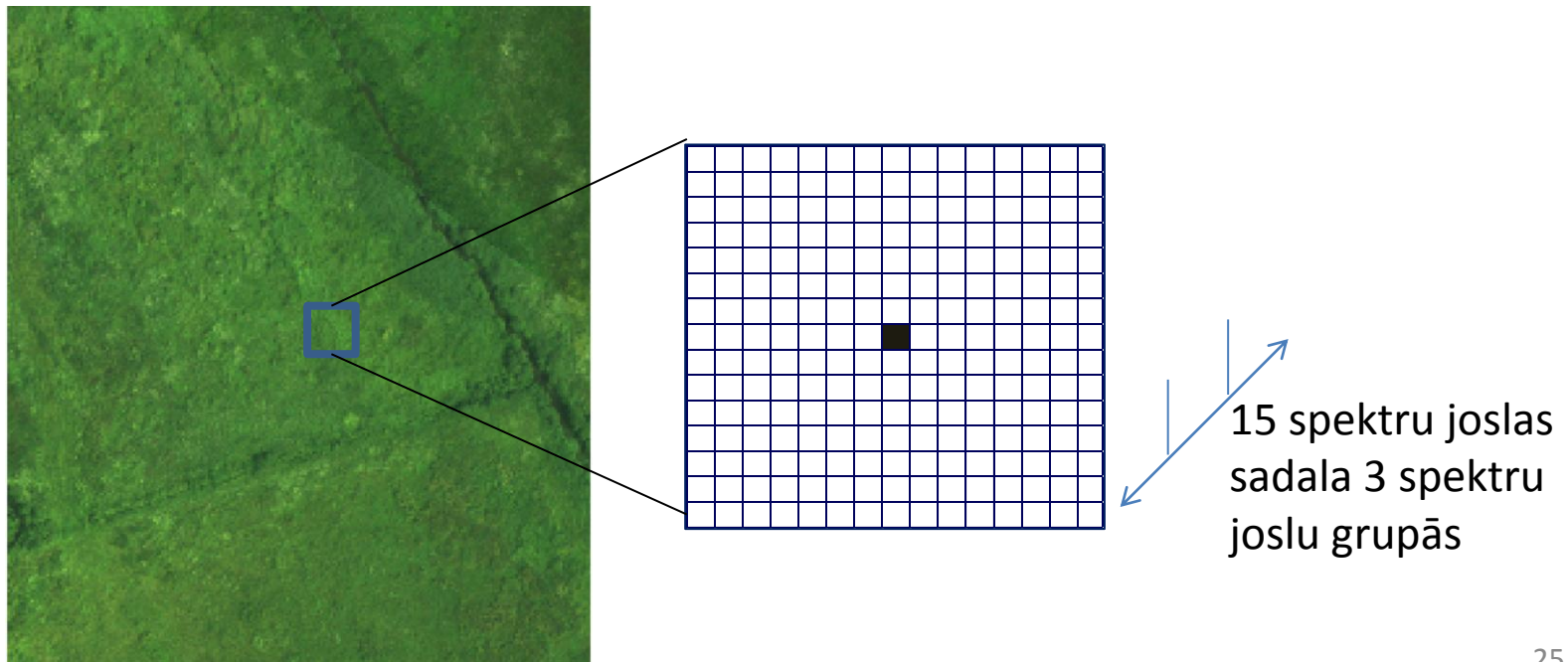
Metode

- **Krāsu dažādību** logā ap pikseli var «mērīt» ar izvēlētā loga pikseļu spektru joslu kovariāciju.
- Tātad ņemt vērā **gan vērtību daudzveidību (izkliedi)** katrā spektrālajā joslā, **gan joslu vērtību kopējo izmaiņu tendencies**.
- Piemēram, ja būtu Red-Green-Blue (RGB) attēls, tas nozīmētu, ka skatās ne tikai tendencies šīm 3 **atsevišķām krāsām**, bet arī šo krāsu pārišu (piem.: sarkana un zaļa) **kompozīciju tendencies**.



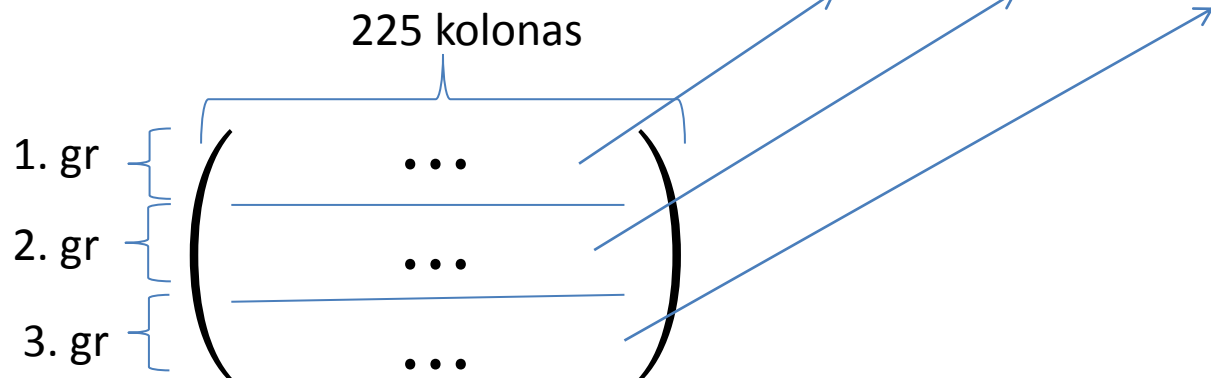
Metode

- Tiek izmantots **logs** 15x15 pikseļi (15m x 15m).
- **Spektru joslas** tika sadalītas **3 informatīvās grupās**.



Konsolidētā kovariācija

- Dati tiek iztīrīti un visām spektra joslām attēlā tiek izlīdzināta skala (veikta **standartizācija**).
- Katram pikselim x_{ij} tiek ņemts **logs (apkārtne)** 15x15 pikseļi.
- No visiem 225 pikseļiem **logā** ap ņemto pikseli x_{ij} katrai no **3 spektru joslu grupām** tiek novērtēta **kovariāciju matrica**: $\hat{\Sigma}_{1ij}, \hat{\Sigma}_{2ij}, \hat{\Sigma}_{3ij}$.



Konsolidētā kovariācija

- Kovariāciju matricu novērtējumiem tiek izskaitļota **Šoleski (Cholesky) faktORIZĀCIJA**, kas ir kvadrātsaknes analogs matricām, ja šīs matricas ir simetriskas un pozitīvi definitas.
- Simetriskai pozitīvi definitai reālai matricai Σ Šoleski faktORIZĀCIJAS rezultātā tiek atrasta apakšējā trijstūra matrica C ar strikti pozitīviem diagonāles elementiem, tāda, ka

$$\Sigma = CC^T.$$

Tātad, no $\hat{\Sigma}_{1ij}$, $\hat{\Sigma}_{2ij}$, $\hat{\Sigma}_{3ij}$ tiek iegūtas $\hat{C}_{1ij}$, $\hat{C}_{2ij}$, $\hat{C}_{3ij}$.

Konsolidētā kovariācija

- Trijstūra matricām, kas iegūtas no **Šoleski faktORIZācijas**, tiek aprēķināta elementu **vidējā vērtība**:

$$u_{1ij} = \frac{1}{n_1} \sum_{c \in \hat{C}_{1ij}} c, \quad u_{2ij} = \frac{1}{n_2} \sum_{c \in \hat{C}_{2ij}} c,$$

$$u_{3ij} = \frac{1}{n_3} \sum_{c \in \hat{C}_{3ij}} c.$$

$$\mathbf{L} = \begin{bmatrix} l_{1,1} & & & & & 0 \\ l_{2,1} & l_{2,2} & & & & \\ l_{3,1} & l_{3,2} & \ddots & & & \\ \vdots & \vdots & \ddots & \ddots & & \\ l_{n,1} & l_{n,2} & \dots & l_{n,n-1} & l_{n,n} \end{bmatrix}$$

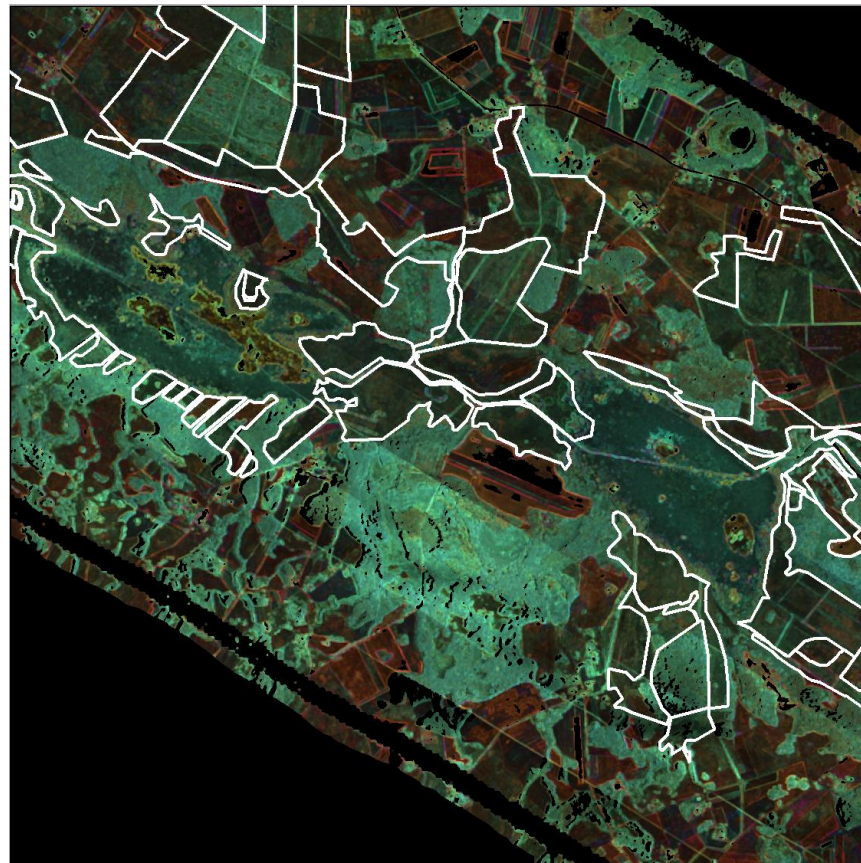
Šeit n_i ir trijstūra matricas elementu skaits.

Konsolidētā kovariācija

- Skaitļi u_{1ij} , u_{2ij} , u_{3ij} piekārtoti katram paņemtam pikselim x_{ij} .
- Vektoru $(u_{1ij}, u_{2ij}, u_{3ij})$ mēs nolēmām saukt par **konsolidēto kovariāciju**.
- Šos skaitļus attēlo kā RGB krāsu, pielāgojot tos krāsu intervālam. Šādu attēlu sauc par **konsolidēto kovariāciju attēlu**.

Krāsu pielāgošana

- Bez krāsu pielāgošanas:



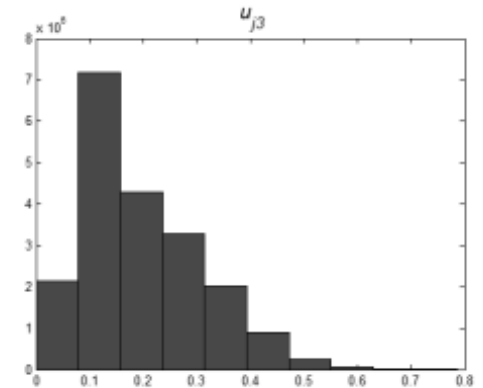
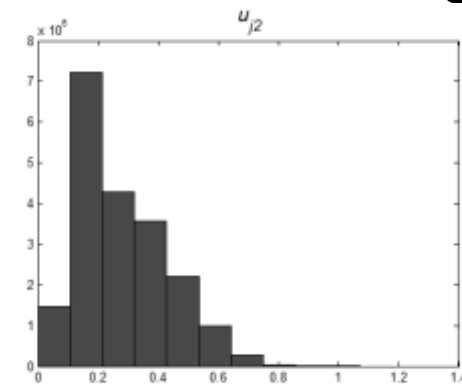
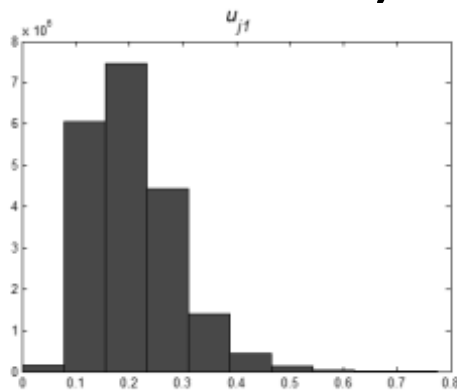
Krāsu pielāgošana

- Ir izmantota **PIT (Probability Integral Transform jeb** «varbūtības integrālā transformācija»**)**, lai «izstieptu» vērtības u_{1ij} , u_{2ij} , u_{3ij} uz intervālu $(0, 1)$.
- Ja ir dots gadījumlielums ξ ar nepārtrauktu sadalījuma funkciju $F_{\xi}(x)$, tad **PIT** pārveido gadījumlielumu ξ par jaunu gadījumlielumu $\eta = F_{\xi}(\xi)$, kura sadalījums, var pierādīt, ir $\text{Uniform}(0,1)$.

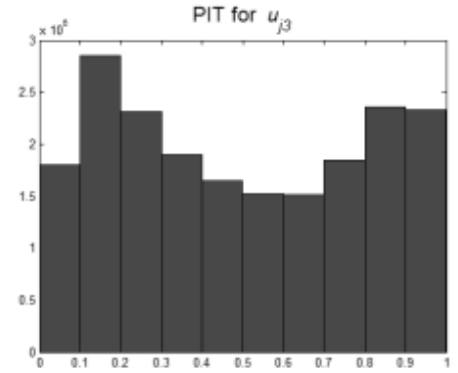
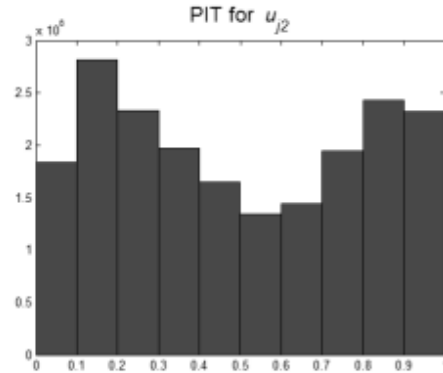
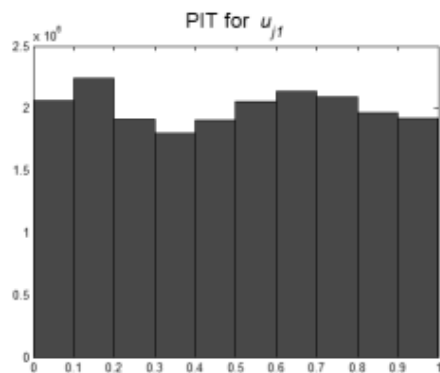
Krāsu pielāgošana

- Tiek atrasts, ka u_{1ij} , u_{2ij} , u_{3ij} sadalījumi ir tuvi gamma sadalījumam (ja tie tiek apskatīti kā iztvērumi). Izveidota attiecīgā PIT.

Bez PIT

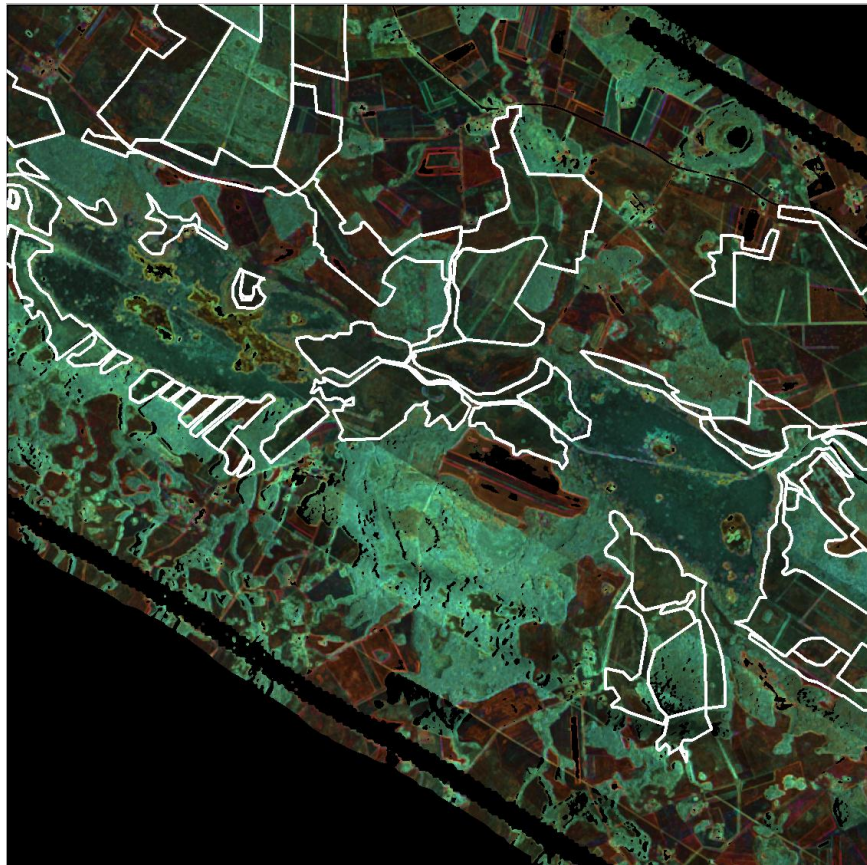


Ar PIT

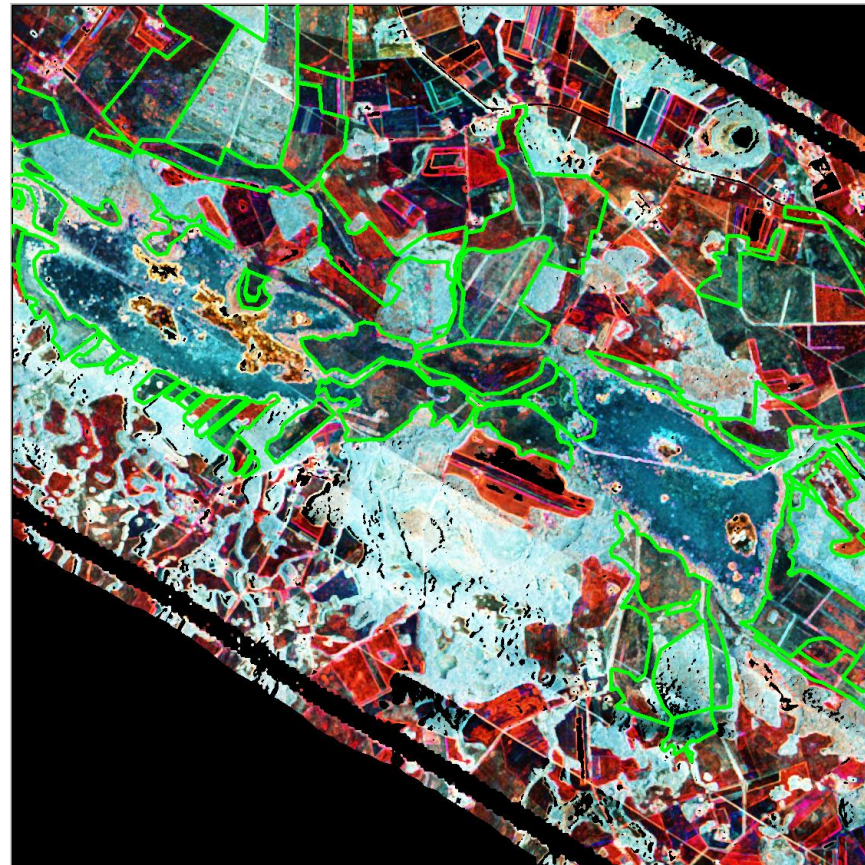


Krāsu pielāgošana

Bez PIT



Ar PIT



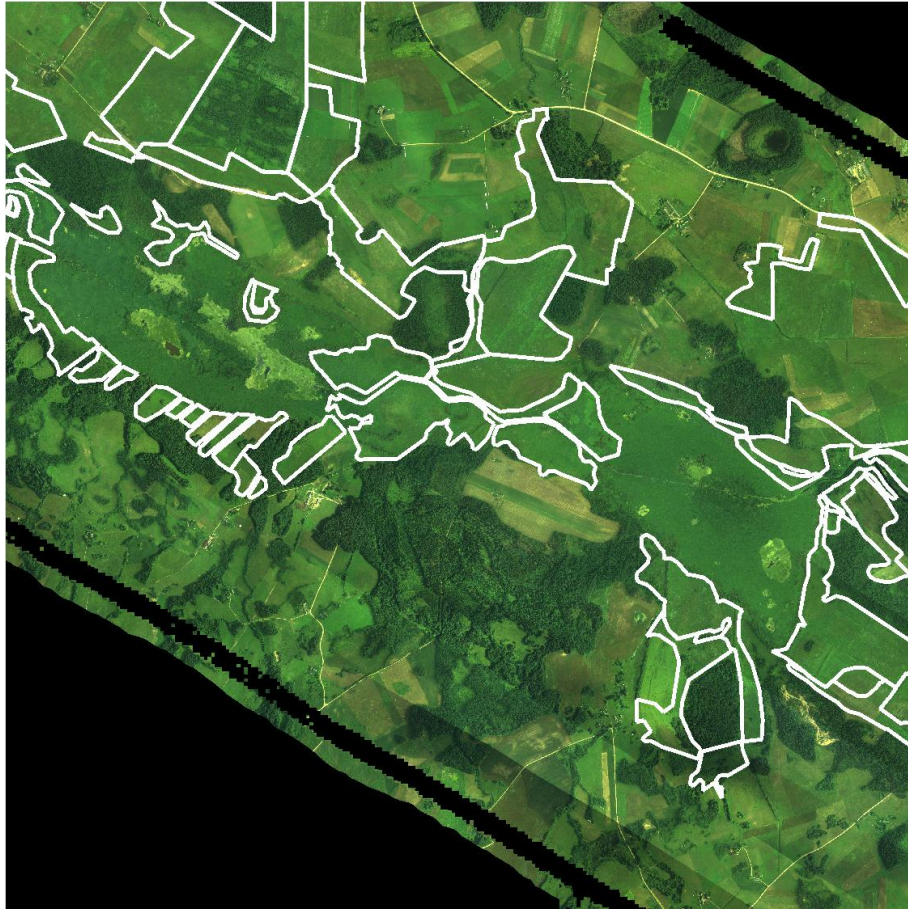
Lieku objektu dzēšana

- Tālāk no attēla tiek izdzēsti lieki objekti.
- Lieko objektu dzēšana no attēla tiek veikta, izmantojot plaknes 3D telpā, kurā atrodas konsolidētās kovariācijas attēla komponentu vērtības.
- Plakņu koeficientus nosaka analītiķis un pārbauda rezultātus vizuāli. Ja rezultāti neapmierina, tad koeficientus pielāgo.
- Ir arī citas metodes, kuras var pielietot.

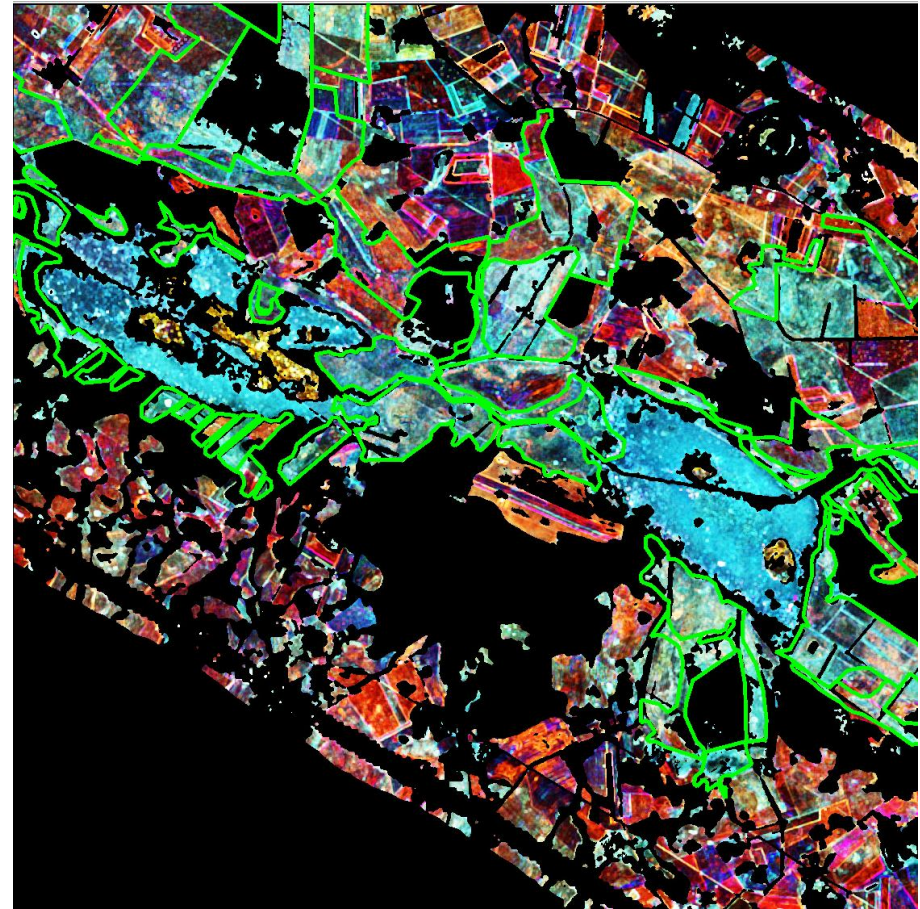
Rezultāti

- Dažādi lauku tipi izdalīti ar nozīmīgi dažādām krāsām, izmantojot **konsolidētās kovariācijas attēlu**.

Parastais RGB attēls

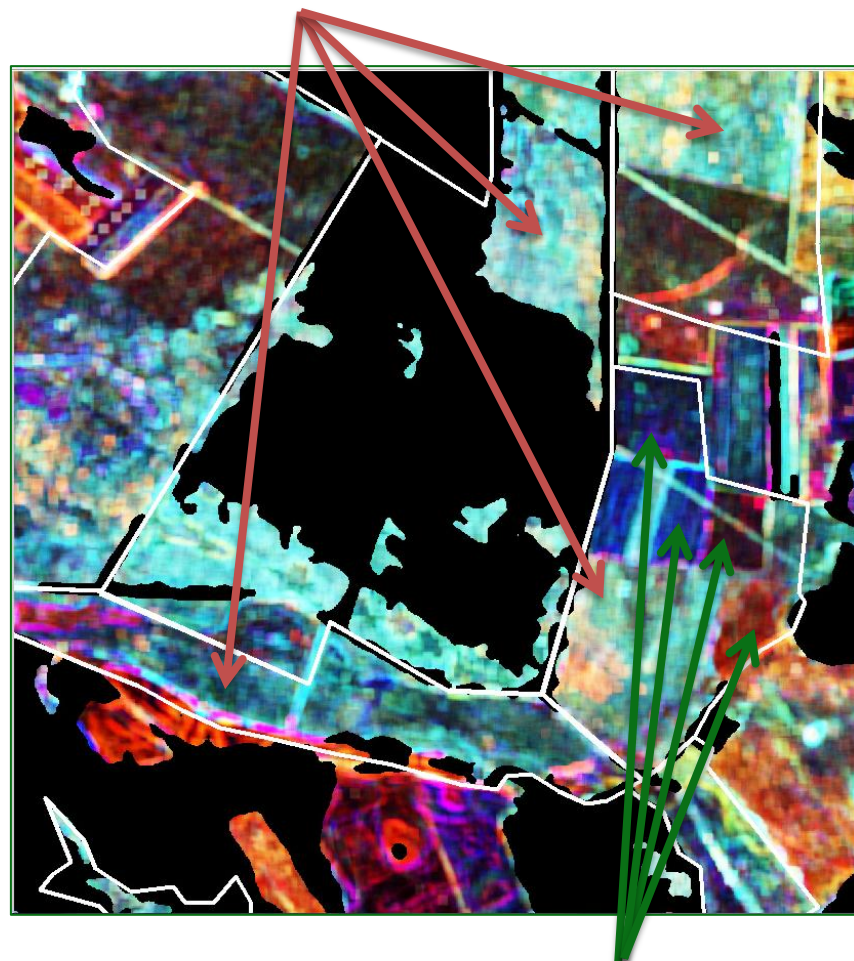


Konsolidētās kovariācijas attēls



Rezultāti

Interesējošiem biotopiem ir raksturīga krāsa.



Ir redzamas vietas, kurās 2005. gadā atzīmētās biotopa robežas (visticamāk) neatbilst 2011. gada datiem.

Secinājumi

- **Konsolidētā kovariācija un konsolidētās kovariācijas attēli varētu būt izmantojami, lai iegūtu papildus informāciju par multispektrālo attēlu.**
- **Izmantojot konsolidēto kovariāciju, interesējošie biotopi varētu būt vieglāk identificējami uz multispektrālā attēla.**
- **Izmaiņas biotopos arī iespējams identificēt, izmantojot konsolidētās kovariācijas attēlu.**
- **Un citi...**

Pielikums: **Andre-Louis Cholesky**



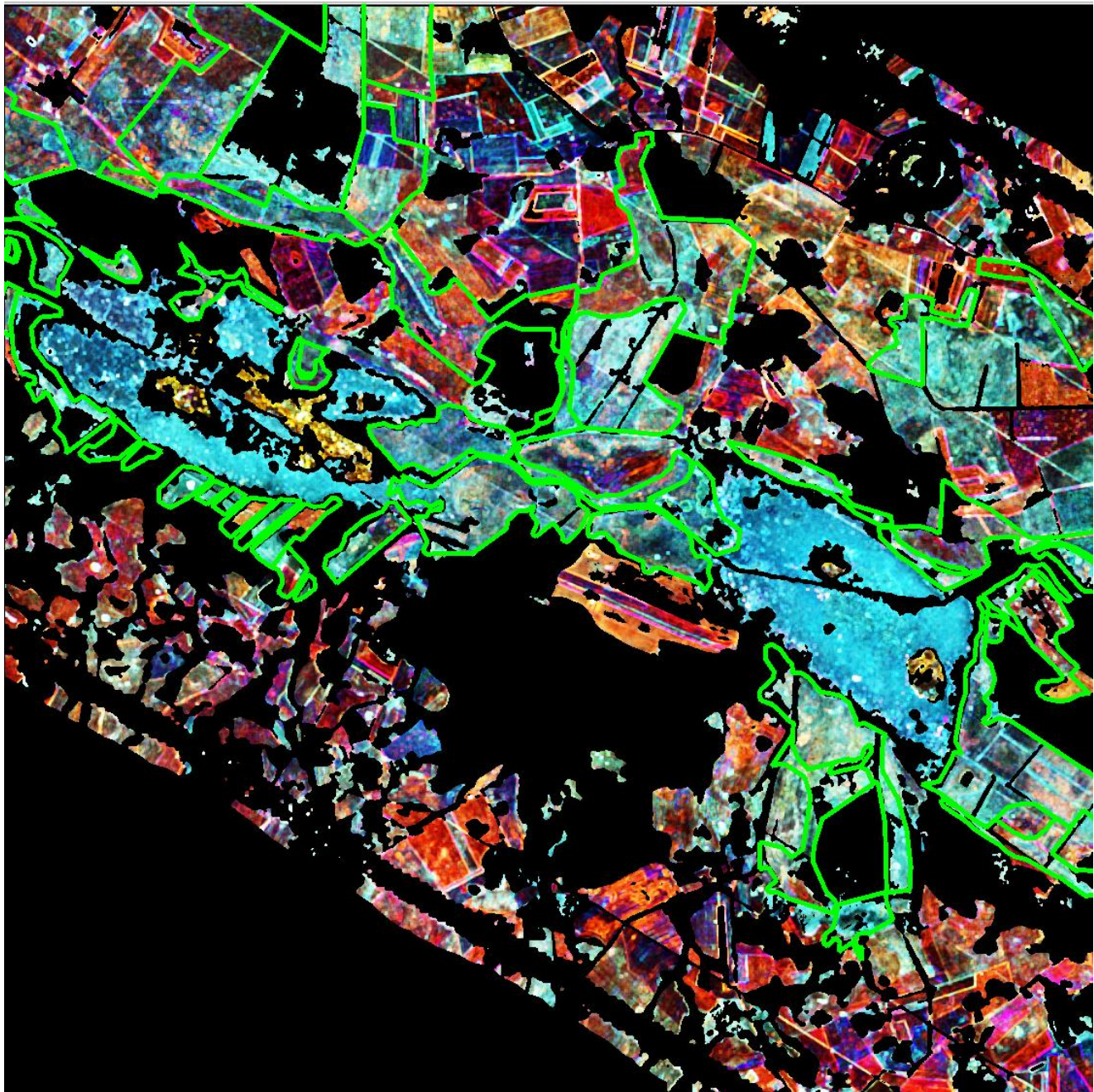
15. Okt. 1875.
(Montguyon, Charentes
Maritime, France) -
31. Aug. 1918. (North
France),
42 gadu vecumā.

- Franču matemātiķis
- Miris no ievainojumiem, iegūtiem kaujas laukā.
- Šoleska faktORIZĀCIJA tiek pielietota:

**Mazāko kvadrātu metodē,
Nelineārā optimizācijā,
Monte Karlo simulācijās,
Kalmana filtros.**

Konsolidētās kovariācijas algoritms

- Izvēlas laukumu logam (mūsu gadījumā 15x15 pikseļi).
- Izlīdzina skalas spektra joslām (standartizē).
- Katram pikselim x_{ij} aprēķina $\hat{\Sigma}_{1ij}, \hat{\Sigma}_{2ij}, \hat{\Sigma}_{3ij}$ (3 grupu gadījumā).
- Izmantojot Šoleski faktORIZāciju, atrod $\hat{C}_{1ij}, \hat{C}_{2ij}, \hat{C}_{3ij}$.
- Atrod šo matricu elementu vidējās vērtības $u_{1ij}, u_{2ij}, u_{3ij}$.
- Ja nepieciešams attēlot vizuāli, tad pielāgo krāsas.
- Uz konsolidētās kovariācijas attēla atrod liekus objektus un izdzēš tos.
- Procedūru atkārt.



Literatūra

- **G.W. Geerling, M. Labrador-Garcia, J.G.P.W. Clevers, A.M.J. Ragas, A.J.M. Smits.** Classification of floodplain vegetation by data-fusion of Spectral (CASI) and LiDAR data. – International Journal of Remote Sensing, Volume 28 Issue 19, September 2007. Taylor & Francis, Inc. Bristol, PA, USA, pp. 4263 – 4284.
- **H. F. Grahn, P. Geladi.** Techniques and Applications of Hyperspectral Image Analysis. – John Wiley & Sons, 2007. 368 p.
- **D. Rocchini, A. Chiarucci, S.A. Loiselle.** Testing the spectral variation hypothesis by using satellite multispectral images. Acta Oecologica 26 (2004), pp. 117 – 120.
- **D. Tiede, S. Lang, F. Albrecht, D. Hölbling.** Class Modeling of Biotope Complexes – Success and Remaining Challenges. ISPRS, Volume XXXVIII-4/C1, 2008 , 6 p.
- **M. Rosenblatt.** Remarks on a Multivariate Transformation. Ann. Math. Statist. Volume 23, Number 3 (1952), pp. 470-472.
- **C. Genest, L.-P. Rivest.** On the multivariate probability integral transformation. Statistics & Probability Letters 53 (2001), pp. 391 – 399

PALDIES PAR UZMANĪBU!



Vinsents Vilems van Gogs. «Zvaigžņotā nakts», 1889. g.