

Kompiuterio mokymo metodai fizikinių ir biocheminių duomenų analizėje

Doktorantas: Tomas Raila

Darbo vadovas: prof. dr. Tadas Meškauskas

2021-09-30

- Studijų laikotarpis: 2017-2021 m.
- Ataskaita už 2020/2021 studijų metus (II pusmetis).

Studijų / ataskaitinio pusmečio planas ir jo vykdymas

Visų studijų planas ir jo vykdymo suvestinė:

| Studijų metai | Egzaminai | | Dalyvavimas konferencijose | | Publikacijos | | |
|-----------------------|-----------|----------|----------------------------|----------|--------------|----------|----------------|
| | Planas | Įvykdyta | Planas | Įvykdyta | Planas | Įvykdyta | Būklė |
| I (2017/2018) | 2 | 2 | | | | | |
| II (2018/2019) | 2 | 2 | 1 | 1 | 1 | 1 | Publikuota |
| III (2019/2020) | | | 1 | 1 | 1 | 1 | Publikuota |
| IV (2020/2021) | | | | | 1 | 1 | Įteikta |

Ataskaitinių metų darbo planas ir jo įvykdymas (2020/2021 studijų metai, II pusmetis):

| Egzaminai | | Dalyvavimas konferencijose | | Publikacijos | |
|-----------|----------|----------------------------|----------|--|---------------------------------|
| Planas | Įvykdyta | Planas | Įvykdyta | Planas | Įvykdyta |
| - | - | - | - | Straipsnis recenzuojamame periodiniame leidinyje, turinčiame citavimo rodiklį Clarivate Analytics Web of Science duomenų bazėje. | Publikacija parengta ir įteikta |

Mokslinių tyrimų ir disertacijos rengimo etapai

| | Darbo pavadinimas | Atlikimo terminai | Pastabos |
|---|---|---|--|
| 1 | <p>Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):</p> <p>1.1. Susipažinti su šiuolaikiniais kompiuterio mokymo metodais (konvoliuciniai, rekurentiniai neuroniniai tinklai, giliojo mokymosi metodai ir t.t.) ir jų taikymais.</p> <p>1.2. Pasirinkti konkrečios dalykinės srities fizikinių ar biocheminių duomenų rinkinį ir susipažinti su jų gavimo procesu.</p> <p>1.3. Apžvelgti pastarųjų metų tyrimus, susijusius su kompiuterio mokymo metodų taikymu pasirinktos dalykinės srities duomenų analizėje.</p> | 2018.04.30 | <ol style="list-style-type: none"> Susipažinta su dalykine sritimi: elektrocheminio impedanso spektroskopija (EIS), atominės jėgos mikroskopija (AFM), baigtiųjų elementų metodu (FEM). Surinkti pradiniai tyrimams reikalingų duomenų rinkiniai. Naudojant baigtiųjų elementų metodą modeliuojamos fosfolipidinių dvisluoksnių membranų su defektais elektrinio laidumo savybės. Modeliai realizuoti naudojant COMSOL Multiphysics paketą, skaičiavimus atliekant MIF paskirstytų skaičiavimų tinkle. Tirta priklausomybė tarp defektų dydžio, tankio, išsibarstymo geometrijos ir membranų elektrinių charakteristikų. Palyginti rezultatai, gauti naudojant sintetiškai sugeneruotus bei realiuose eksperimentuose gautus defektų pasiskirstymus. Pradėtas kurti metodas automatiniam defektų aptikimui atominės jėgos mikroskopo nuotraukose. |
| 2 | <p>Mokslinio tyrimo vykdymas:</p> <p>2.1. Tyrimo metodikos sudarymas:</p> <p>2.1.1. Sukurti duomenų paruošimo modelių apmokymui metodiką.</p> <p>2.1.2. Sudaryti fizikinių ir biocheminių duomenų analizės algoritmus, paremtus kompiuterio mokymo metodais.</p> <p>2.1.3. Parašyti arba iš egzistuojančių programinės įrangos paketų pasirinkti programas, realizuojančias tyrimu numatytus kompiuterio mokymo modelius.</p> <p>2.2. Teorinis tyrimas:</p> <p>2.2.1. Nustatyti parinktų kompiuterio mokymo algoritmų pradinis parametras, atlikti statistinę duomenų rinkinio analizę</p> <p>2.2.2. Paruošti apmokymo duomenų aibę, parengti sintetinių duomenų generavimo metodiką.</p> <p>2.2.3. Remiantis sudarytais matematiniais ir kompiuteriniais modeliais numatyti naujus rezultatus.</p> <p>2.3. Empirinis tyrimas:</p> <p>2.3.1. Sugeneruoti sintetinių duomenų rinkinius, palyginamus su realiais duomenimis.</p> <p>2.3.2. Atlikti modelių apmokymą su sintetinių ir realių duomenų rinkiniais, naudojant paskirstytų skaičiavimų resursus ar specializuotą aparatinę įrangą (grafinius</p> | <p>2018.10.30</p> <p>2019.10.30</p> <p>2020.01.30</p> | <ol style="list-style-type: none"> Sudaryti ir eksperimentiškai iširti du teoriniai defektų klasterizacijos modeliai, lyginant sintetiškai sugeneruotus defektų pasiskirstymus su eksperimentiškai užregistruotais. Pasiūlyta metodika, leidžianti kiekybiškai išreikšti defektų klasterizacijos stiprumą pagal erdvinį defektų išsidėstymą membranoje. Naudojant baigtiųjų elementų metodą modeliuotos membranos su skirtingais defektų klasterizacijos modeliais bei skirtingais jų parametrais. Tiriamas ryšys tarp realiose membranose esančių defektų pasiskirstymų ir teorinių klasterizacijos modelių parametrų. Pasiūlyta pradinė metodika, leidžianti kokybiškai įvertinti defektų klasterizacijos egzistavimą iš eksperimentinio EIS spektro. Tirtos FEM modeliavimo proceso optimizacijos galimybės, parenkant optimalų tinklėlio tankį, tiesinių lygčių sistemų sprendimo metodą bei jo parametrus ir skaičiavimų aplinką. Sprendžiamas defektų aptikimo AFM mikroskopijos nuotraukose uždavinys, taikant nesudėtingus skaitmeninių vaizdų analizės metodus (Hough transformacija). |

Mokslinių tyrimų ir disertacijos rengimo etapai (2)

| | | | |
|---|--|------------|---|
| | <p>procesorius).</p> <p>2.3.3. Verifikuoti skaičiavimų rezultatus, palyginant juos su ekspertiniu vertinimu bei literatūroje pateiktais vertinimais.</p> <p>2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas:</p> <p>2.4.1. Palyginti su skirtingais modeliais gautus rezultatus tarpusavyje bei kitų autorių tyrimų kontekste. Išskirti efektyviausius modelius bei jų parametrus, optimizuoti jų taikymą.</p> | 2020.10.30 | |
| 3 | <p>Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:</p> <p>1. Pasiūlyti naują ar patobulintą kompiuterio mokymo metodiką, tinkamą pasirinktos dalykinės srities duomenų analizei.</p> <p>2. Iširti pasiūlytų metodų optimizavimo, atsižvelgiant į taikymus fizikoje ir biochemijoje, galimybes.</p> <p>3. Pateikti duomenų analizės, atliktos pasiūlytais kompiuterio mokymo metodais, rezultatus, juos palyginti su kitų autorių publikuotais rezultatais.</p> | 2021.02.01 | <ol style="list-style-type: none"> 1. Taikant mašininio mokymosi metodus sprendžiamas defektų aptikimo AFM vaizduose uždavinys. Gauti pradiniai rezultatai naudojant konvoliucinį neuroninį tinklą. 2. Lyginami eksperimentiškai gauti membranų EIS spektrai ir jų atitikmenys, gauti atlikus baigtinių elementų modeliavimą pagal žinomas arba AFM vaizduose atpažintas defektų koordinates. 3. Tiriamas ryšys tarp defektų aptikimo AFM vaizduose tikslumo bei pokyčių sumodeliuotuose EIS spektruose, naudojant sintetiškai sugeneruotus defektų pasiskirstymus, atitinkančius įvairaus tikslumo defektų aptikimo rezultatus. 4. Pradėtas rengti disertacijos tekstas. |
| 4 | Daktaro disertacijos parengimas ir svarstymas padalinyje | 2021.06.30 | <ol style="list-style-type: none"> 1. Pritaikyti alternatyvūs defektų aptikimo AFM vaizduose metodai (Hough transformacija, šablono atitikimas), palygintas jų tikslumas su konvoliucinio neuroninio tinklo rezultatais. 2. Kiekybiškai įvertinta defektų aptikimo AFM vaizduose tikslumo įtaka modeliuojamiems EIS spektrams, pastarieji palyginti su eksperimentiškai gautais EIS duomenimis. 3. Apibrėžta galutinė disertacijos struktūra, pagrindinės dalys pildomos remiantis publikacijų medžiaga. |
| 5 | Daktaro disertacijos gynimas | 2021.09.30 | Parengtas pirminis disertacijos variantas (juodraštis). |

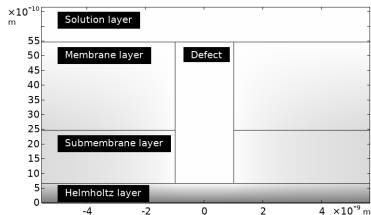
- Raila T., Penkauskas T., Jankunec M., Dreižas G., Meškauskas T., Valinčius G., (2019). Electrochemical impedance of randomly distributed defects in tethered phospholipid bilayers: Finite element analysis.
 - 2019 m. publikuotas žurnale *Electrochimica Acta*
- Raila T., Ambrulevičius F., Penkauskas T., Jankunec M., Meškauskas T., Vanderah D. J., Valinčius G., (2020). Clusters of protein pores in phospholipid bilayer membranes can be identified and characterized by electrochemical impedance spectroscopy.
 - 2020 m. publikuotas žurnale *Electrochimica Acta*
- Raila T., Penkauskas T., Ambrulevičius F., Jankunec M., Meškauskas T., Valinčius G., (2021). AI-based Atomic Force Microscopy Image Analysis Allows to Predict Electrochemical Impedance Spectra of Tethered Bilayers Affected by Pore-Forming Toxins.
 - **Publikacija parengta ir įteikta**

- Raila. T., Meškauskas T., Valinčius G., Jankunec M., Penkauskas T. (2019) Computer modeling of electrochemical impedance spectra for defected phospholipid membranes: finite element analysis
 - 2020 m. publikuotas Springer LNCS 11974
- Raila T., Jankunec M., Meškauskas T., Valinčius G., (2020). Computational models of defect clustering for tethered bilayer membranes.
 - 2020 m. publikuotas Springer LNCS 12253

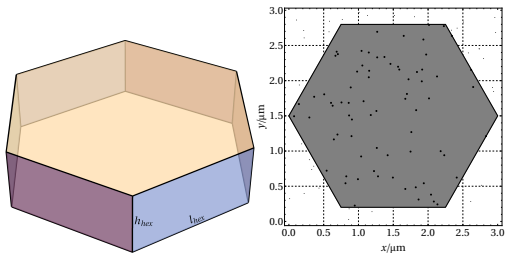
- 10-oji tarptautinė konferencija „Data Analysis Methods for Software Systems“ (DAMSS2018), 2018-11-28 – 2018-11-30, Druskininkai, Lietuva.
- 9-oji Jaunųjų mokslininkų konferencija „Fizinių ir technologijos mokslų tarpdalykiniai taikymai“. 2019-03-12, Vilnius, Lietuva.
- 3-oji tarptautinė konferencija „Numerical Computations: Theory and Algorithms“ (NUMTA2019), 2019-06-15 – 2019-06-20, Krotonė, Italija.
- 20-oji tarptautinė konferencija „International Conference on Computational Science and Its Applications“ (ICCSA2020), 2020-07-01 – 2020-07-04, Kaljaris, Italija.
 - Dalyvauta nuotoliniu būdu

- Nagrinėjami trimačiai prikabintų dvisluoksnių lipidinių membranų (tBLM) modeliai su laisvai apibrėžtais defektų išsidėstymais.
 - Bendradarbiaujama su VU GMC Biochemijos instituto mokslininkais.
- Baigtinių elementų metodu (FEM) modeliuojamos EIS (elektrocheminio impedanso spektroskopija) spektrų savybės ir tiriamas jų ryšys su defektų pasiskirstymo tipu, jų tankiu bei dydžiu.
- Analizuojami membranų skaitmeniniai vaizdai, gauti naudojant atominės jėgos mikroskopiją (AFM).

Trimatis membranos modelis



(a) Trimačio membranos modelio skerspjūvis defekto aplinkoje



(b) Modelio geometrija (c) Defektų išsidėstymas

- FEM sprendžiama Laplaso lygtis:

$$\nabla \cdot (\tilde{\sigma}(x, y, z) \nabla \Phi(x, y, z)) = 0 \quad (1)$$

$$\tilde{\sigma}(x, y, z) = \sigma(x, y, z) + j \omega \varepsilon(x, y, z), \quad (2)$$

- Kraštinės sąlygos:

$$\Phi(x, y, h_{hex}) = 1 \quad (3)$$

$$\Phi(x, y, 0) = 0 \quad (4)$$

$$n \cdot J = 0 \quad (5)$$

- Srovės tankis:

$$J(x, y, z) = -\tilde{\sigma}(x, y, z) \nabla \Phi(x, y, z) \quad (6)$$

- Admitansas:

$$Y = \frac{\iint_{(x,y) \in \Gamma_{hex}} -n \cdot J(x, y, h_{hex}) dx dy}{S_{hex}} \times \frac{1}{\Phi(x, y, h_{hex})} \quad (7)$$

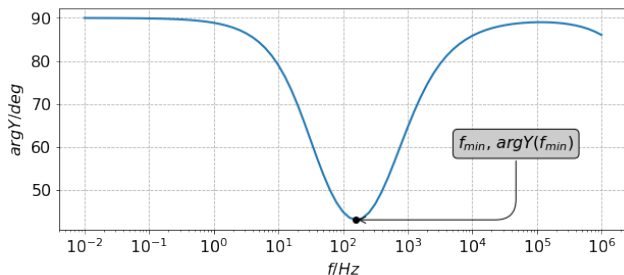
EIS parametrai

- EIS - elektrocheminio impedanso spektroskopija
- Išvestiniai parametrai:

f_{min} – dažnis f kuriame $\arg Y(f)$ įgyja mažiausią reikšmę, (8)

$\arg Y(f_{min})$ – admitanso fazės reikšmė taške f_{min} , (9)

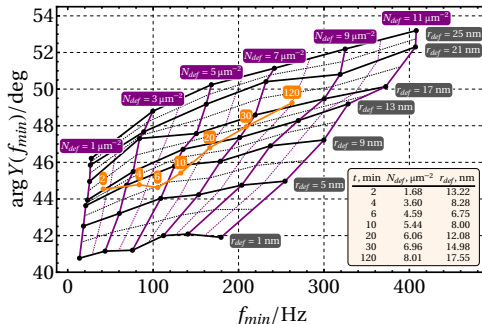
(10)



2 pav.: Sumodeliuoto EIS spektro pavyzdys

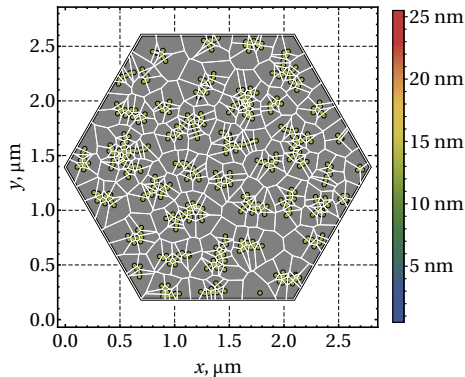
Membranos parametry kokybinis įvertinimas

- Egzistuoja kokybinis panašumas tarp analitiškai gautų EIS spektrų homogeniškiems defektų pasiskirstymams bei skaitiškai sumodeliuotų spektrų esant atsitiktiniams pasiskirstymams.
- Atsitiktinai kompiuteriu sugeneruotų ir eksperimentiškai gautų defektų pasiskirstymų (esant fiksuotam tankiui) EIS spektrai sutampa tik pasiskirstymų be klasterių atvejais.
- Pagal eksperimentiškai gautus EIS spektrus įmanoma įvertinti membranos defektų savybes.

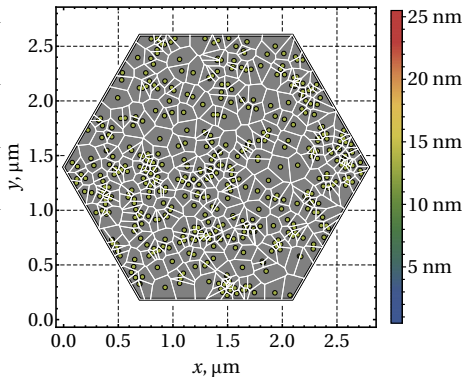


Defektų klasterizacijos modeliai

- Atsitiktinis išsidėstymas (klasterizacijos nėra)
- Defektų tarpusavio trauka (3a pav.)
- Procedūrinis triukšmo generavimas (*lattice convolution noise - LCN*) (3b pav.)



(a) Defektų tarpusavio trauka



(b) Procedūrinis triukšmo generavimas

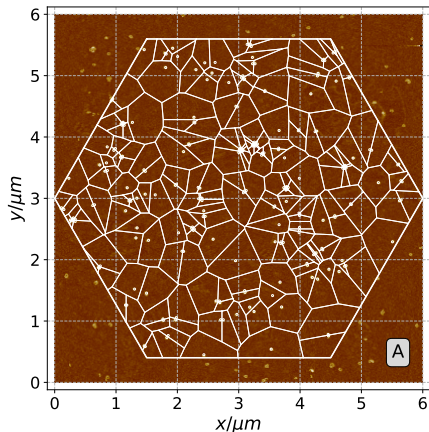
Klasterizacijos rodikliai

σ - Voronojaus diagramos sektorių plotų standartinis nuokrypis

- Atsitiktinis defektų išsidėstymas: $\sigma \approx 0.54$
- Realūs defektų išsidėstymai, fiksuoti AFM: $\sigma > 0.8$

Kitos metrikos:

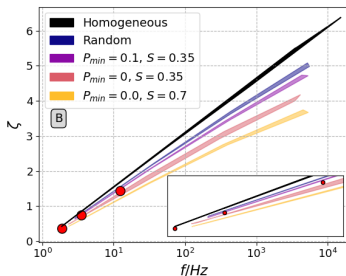
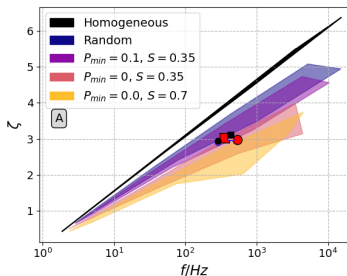
- Asimetrijos koeficientas (skewness)
- Ekscesas (kurtosis)
- Absoliutusis medianos nuokrypis (MAD)



4 pav.: Membranos AFM vaizdas su defektų Voronojaus diagrama

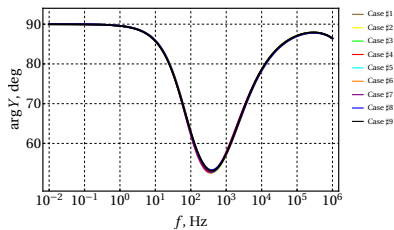
Klasterizācijas kokybinis jvertinimas

- Defektu išsidėstymo pobūdžio identifikavimas: homogeniškas (reguliarus), atsitiktinis ar klasterizuotas.
- Klasterizacijos jvertinimas pagal spektro ruožo aplink minimumą formą, pasiūlyta metrika ζ

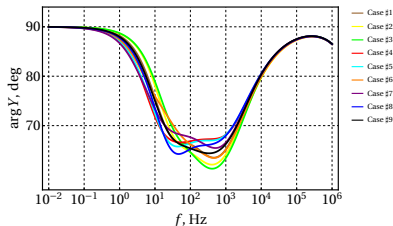


Klasterizacijos kiekybinis įvertinimas

- Pasireiškiant defektų klasterizacijai keičiasi EIS spektrų forma, atsiranda papildomi minimumai.
- Klasterizacijos efekto įvertinimas pagal EIS spektrą:
 - Voronojaus diagramos rodiklis σ
 - LCN ar kito modelio parametrai

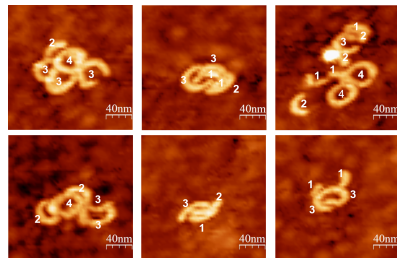


(a) Atsitiktinis defektų išsidėstymas



(b) Klasterizuotas defektų išsidėstymas (LCN modelis)

- Duomenų rinkinys: po 9 vaizdų fragmentus iš 3 membranų paviršių.
 - Vienas fragmentas (512 x 512 raiška) atitinka $2\mu m \times 2\mu m$ paviršiaus plotą
- Tikrosios defektų koordinatės sužymėtos eksperto (apie 1000 pavyzdžių)
- Tiems patiems membranų bandiniams papildomai atlikti EIS matavimai

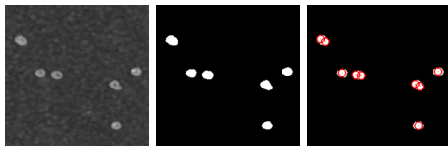


6 pav.: Defektų klasteriai

Defektų aptikimas AFM vaizduose

- Defektų zonų plotų matavimas
 - Slenkstinis filtras (angl. threshold)
 - Morfologinis apdorojimas
 - Zonos dydis vaizdo taškais / vidutinis vieno defekto dydis = defektų kiekis zonoje
 - Defektų koordinatės priskiriamos klasterizuojant zonos taškus su K-means
- Rezultatai:

| AFM surface | N_{true} | N_{pred} | Precision | Recall | F1 |
|-------------|------------|------------|-----------|--------|-------|
| 1 | 172 | 114 | 0.860 | 0.570 | 0.685 |
| 2 | 97 | 114 | 0.553 | 0.649 | 0.597 |
| 3 | 158 | 194 | 0.613 | 0.753 | 0.676 |



7 pav.: AFM vaizdo apdorojimo žingsniai

Defektų aptikimas AFM vaizduose (2)

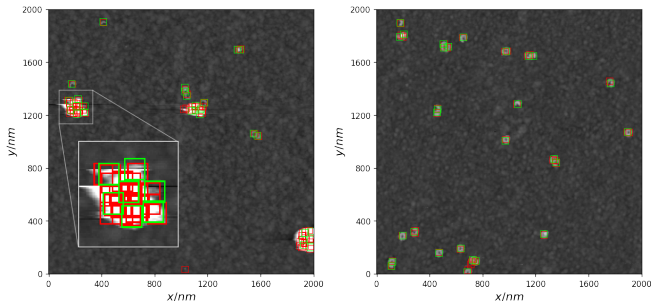
- Hough transformacija
 - Slenkstinis filtras
 - Morfologinis apdorojimas
 - Apskritiminė Hough transformacija (CHT)
- Parametrai: Hough slenkstis, galimų apskritimų spindulių intervalas.
- Rezultatai:

| AFM surface | N_{true} | N_{pred} | Precision | Recall | F1 |
|-------------|------------|------------|-----------|--------|-------|
| 1 | 172 | 201 | 0.577 | 0.674 | 0.622 |
| 2 | 97 | 156 | 0.436 | 0.701 | 0.538 |
| 3 | 158 | 140 | 0.614 | 0.544 | 0.577 |

Defektų aptikimas AFM vaizduose (3)

- Konvoliucinis neuroninis tinklas (CNN)
 - Architektūra: SSD FPN ResNet-50
 - Bazinis modelis apmokytas su COCO vaizdų rinkiniu
 - Pritaikytas defektų aptikimui AFM vaizduose
 - Apmokymo aibė - 15 vaizdų fragmentų, testavimo aibė - 12 fragmentų

| AFM surface | Precision | Recall | F1 |
|-------------|-----------|--------|-------|
| 1 | 0.775 | 0.581 | 0.664 |
| 2 | 0.555 | 0.680 | 0.611 |
| 3 | 0.757 | 0.728 | 0.742 |

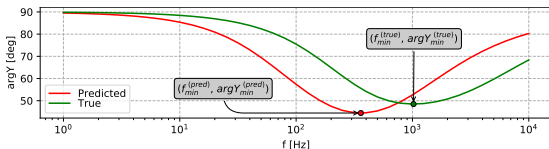


Defektų aptikimo tikslumas ir pokyčiai EIS spektruose

- Tikslumo vertinimo metrikos: *precision*, *recall*, F1
- EIS spektrų, atitinkančių tikrąsias ir prognozuotas defektų koordinates, palyginimas minimumo taško atžvilgiu:

$$\Delta f_{log} = \log_{10}(f_{min}^{(true)}) - \log_{10}(f_{min}^{(pred)})$$

$$\Delta \arg Y = \arg Y_{min}^{(true)} - \arg Y_{min}^{(pred)}$$

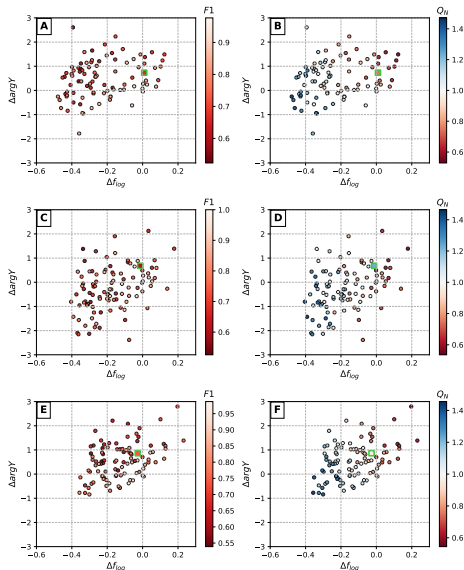


- Pasirinkus tikrų (eksperto sužymėtų) defektų koordinacių rinkinį, iš jo generuojami sintetiniai rinkiniai, pasižymintys skirtingomis $F1$ ir Q_N reikšmėmis:

$$Q_N = N_{def}^{(pred)} / N_{def}^{(true)}$$

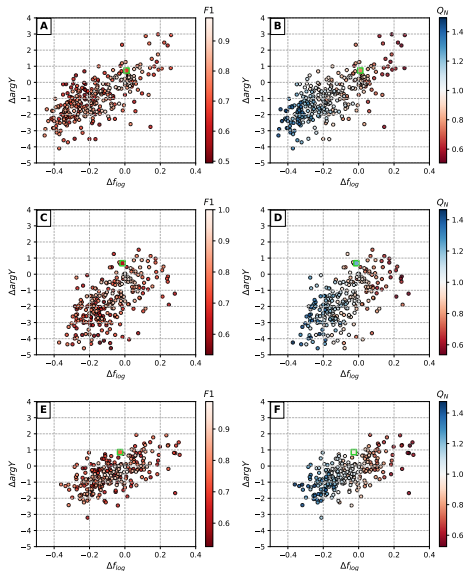
Neklasterizuotų sintetinių def. rinkinių generavimas

- Esami defektai pastumiami prie koordinatinių pridodant atsitiktines reikšmes
- Dalis atsitiktinai parinktų defektų pašalinama (imituojami klaidingai neigiami atvejai)
- Pridedama dalis defektų su atsitiktinai parinktomis koordinatėmis (imituojami klaidingai teigiami atvejai)
- **Problema:** sisteminiai nuokrypiai sąlygojami kintančios klasterizacijos



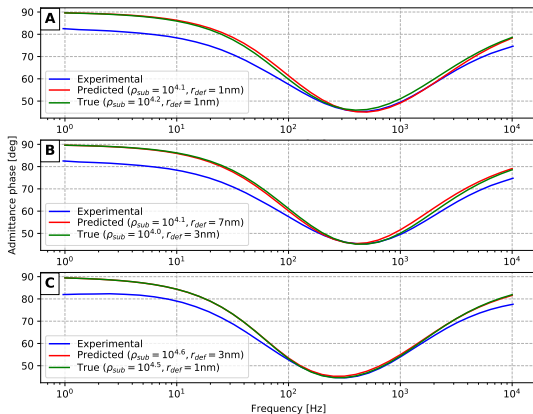
Klasterizuotų sintetinių defektų rinkinių generavimas

- Pagal tikrąsias defektų koordinatas sudaromas KDE (angl. kernel density estimation) modelis
- Esami defektai pastumiami prie koordinatinių pridant atsitiktines reikšmes
- Pašalinama dalis defektų esančių arčiausiai atsitiktinai parinktų taškų (iš KDE modelio)
- Pridedama dalis naujų defektų su atsitiktinai parinktomis koordinatėmis iš KDE modelio



Palyginimas su eksperimentiniais EIS duomenimis

- EIS spektrai sumodeliuoti su tikromis ir prognozuotomis defektų koordinatėmis sugretinti su eksperimentinėmis kreivėmis
- Būdas apytiksliai įvertinti parametrą ρ_{sub} , kuris nėra išmatuojamas eksperimentiškai.



Ačiū už dėmesį!