

# Atsitiktinių medžių (*random forest*) klasifikatoriaus taikymas snygio nustatymui iš palydovinių ATMS duomenų

## Application of Random forest Algorithm to Detect Snowfall from ATMS Measurements

Justinas KILPYS<sup>1</sup>, Huan MENG<sup>2</sup>, Yalei YOU<sup>3</sup>, Jun DONG<sup>3</sup>, Ralph FERRARO<sup>2</sup>

<sup>1</sup>Vilniaus universitetas, Chemijos ir geomokslų fakultetas, Geomokslų institutas, justinas.kilpys@gf.vu.lt

<sup>1</sup>Vilnius University, Faculty of Chemistry and Geosciences, Institute of Geosciences, justinas.kilpys@gf.vu.lt

<sup>2</sup>NOAA/NESDIS/Center for Satellite Applications and Research (STAR), huan.meng@noaa.gov, ralph.r.ferraro@noaa.gov

<sup>3</sup>University of Maryland/ESSIC/Cooperative Institute for Satellite Earth System Studies (CISESS),  
yyou@umd.edu, jundong@umd.edu

DOI: <https://doi.org/10.15388/Klimatokaita.2020.37>

Kietų kritulių nustatymas iš dirbtinių Žemės palydovų, naudojant pasyvius mikrobangų prietaisus, pasižymi didesnėmis paklaidomis nei nustatant skystus kritulius. Šiame tyrime buvo siekiama nustatyti snygio atvejus panaudojant palydovinius ATMS (*Advanced Technology Microwave Sounder*) ryškio temperatūros duomenis. Snygio aptikimui buvo panaudotas atsitiktinių medžių (angl. *Random Forest*) klasifikatorius. Klasifikatoriui sudaryti ir jo tikslumui įvertinti naudoti 2015–2018 m. globalūs meteorologinių stočių, ATMS prietaiso bei NOAA GFS modelio (*Global Forecast System*) duomenys. Naudojant surinktus duomenis sudarytas hibridinis snygio nustatymo algoritmas. Palydoviniai ir orų modelio duomenys naudoti kaip prediktoriai, o antžeminiai meteorologinių stočių kritulių stebėjimai – kaip pirminiai duomenys, skirti apmokyti klasifikatorių.

Svarbiausi ATMS skenavimo kanalai ir GFS modelio rodikliai buvo atrinkti naudojant atsitiktinių medžių klasifikatoriaus kintamųjų reikšmingumo rodiklį. Nustatyta, kad geriausi snygio prediktoriai yra aukšto dažnio ATMS kanalai 17, 18 ir 19 (165–183 GHz). Snygio aptikimo tikslumas padidėjo, kai į atsitiktinių medžių algoritmą buvo įtraukti GFS modelio santykinės drėgmės ir oro temperatūros 800–900 hPa lygio duomenys.

Atsitiktinių medžių klasifikatoriaus vertinimas parodė, kad, esant labai šaltam orui (kai temperatūra 2 m aukštyje žemesnė negu  $-10$  °C), ATMS duomenys nepadedą atskirti snygio nuo atvejų be kritulių. Esant tokioms sąlygoms, geresni prediktoriai buvo GFS bendras vandens garų kiekis atmosferoje (CWAT) ir oro temperatūra apatinėje troposferoje (800–900 hPa).

Nors snygio aptikimas naudojant ATMS ryškio temperatūrą vis dar kelia sunkumų (ypač labai šaltu oru), atsitiktinių medžių klasifikatoriaus panaudojimas leido pasiekti geresnių rezultatų nei naudojant įprastinę tiesinę logistinę regresiją.

Tyrimas atliktas 2019 m. mokslinės stažuotės CISESS, kurią finansavo Baltijos–Amerikos laisvės fondas, metu.