



Universiteit
Leiden

The Netherlands

The state of the earth: estimating physical parameters from noisy and incomplete earth observation data

Arp, L.R.

Citation

Arp, L. R. (2026, June 23). *The state of the earth: estimating physical parameters from noisy and incomplete earth observation data*. Retrieved from <https://hdl.handle.net/1887/4306907>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4306907>

Note: To cite this publication please use the final published version (if applicable).

SAMENVATTING

Het inschatten van wetenschappelijke parameters op basis van observaties is een centraal thema in wetenschappelijke AI, met brede relevantie voor diverse wetenschappelijke vakgebieden en applicaties. Deze parameters beschrijven de status van een fysiek systeem, en zijn vaak niet direct te observeren: we kunnen ze alleen inschatten via geobserveerde uitkomsten. In dit proefschrift ligt de focus op het ontwikkelen van betrouwbare AI-oplossingen om aardwetenschappelijke parameters in te schatten met behulp van satellietdata. Deze parameters zijn essentieel voor, onder andere, milieubescherming, rampenbestrijding en landbouw. Er zijn echter twee centrale uitdagingen die deze inschatting bemoeilijken: 1) in onze data ontbreken grote hoeveelheden spatiëel-temporele informatie, waardoor het trainen van modellen wordt bemoeilijkt en de consistentie van voorspellingen wordt verminderd, en 2) meerdere verschillende fysieke omstandigheden kunnen dezelfde observaties genereren, waardoor er meerdere goed passende oplossingen zijn voor een observatie. In dit proefschrift beantwoorden we vier onderzoeksvragen, waarbij er twee gewijd zijn aan de eerste uitdaging, en twee aan de tweede uitdaging.

De eerste onderzoeksvraag betreft ontbrekende informatie in spatiëel-temporele ground truth datasets, en hoe deze op een effectieve manier ingevuld kunnen worden. In Hoofdstuk 3 stellen we een nieuwe spatiële interpolatiemethode voor, om zo de invloed van dit probleem te verminderen. Ground truth data, die de ‘echte’ waarden bevat voor de parameters die we moeten inschatten, beslaat vaak slechts een groep specifieke meetpunten verspreid over een onderzoeksgebied, terwijl we meestal geïnteresseerd zijn in een volledige kaart met voorspellingen op alle punten in het gebied. Onze methode, geïnspireerd door Markov reward processes, propageert iteratief de beschikbare informatie door een systeem van onbekende waarden, met als eindresultaat een geïnterpoleerd raster met ground truth waarden. We tonen de effectiviteit van onze methode aan op zowel gesimuleerde data als echte BBP en COVID-19 datasets.

De tweede onderzoeksvraag betreft ontbrekende spatiëel-temporele informatie in satellietdata, bijvoorbeeld veroorzaakt door wolkendekking. In Hoofdstuk 4 stellen we een nieuwe methode voor die onze interpolatiemethode van Hoofdstuk 3 aanpast voor het reconstrueren van afbeeldingen. Onze nieuwe methode stuurt het interpolatiealgoritme via een volledig geobserveerde maar verouderde referentie-afbeelding, en vult hiermee de missende data in. We tonen de effectivi-

teit van deze methode aan op zowel een populaire dataset voor het verwijderen van wolkendekking, als op een dataset met hoge diversiteit die wij zelf hebben gecreëerd. Wij maken deze dataset publiek toegankelijk voor verder wetenschappelijk onderzoek.

De derde onderzoeksvraag betreft de meerdere passende oplossingen bij het inschatten van parameters. Bij het gebruik van een veel gebruikte inversietechniek voor radiatieve transfer models (RTMs) kwamen we onverwacht tot de ontdekking dat er slechts een enkele oplossing mogelijk was, in tegenstelling tot onze verwachtingen. In Hoofdstuk 5 voeren we een empirisch onderzoek uit op data gesimuleerd door een RTM, om te begrijpen hoe problemen met een enkele oplossing zich in de praktijk toch als meerdere passende oplossingen kunnen presenteren. Uit onze analyse komt naar buiten dat het probleem alle criteria van een well-posed probleem (namelijk dat er een unieke oplossing bestaat), en experimenten op echte satellietdata bevestigen dat deze eigenschap behouden blijft buiten een puur gesimuleerde context. Uit verdere experimenten blijkt dat ruis op de observaties waarschijnlijk de voornaamste oorzaak is van meerdere passende oplossingen. Zelfs als er een unieke oplossing bestaat voor een observatie, leiden fouten in de observatie zelf tot fouten in de oplossingen.

De laatste onderzoeksvraag betreft het berekenen van de set van passende oplossingen voor het inschatten van parameters. In Hoofdstuk 6 formaliseren we dit probleem, en geven we een theoretische onderbouwing voor een kader om deze berekening computationeel haalbaar te maken. We stellen ook een methode voor, gebaseerd op constrained black-box optimalisatie, om automatisch de set van mogelijke oplossingen te benaderen. We valideren zowel het theoretische kader als onze praktische benaderingsmethode via empirische experimenten met natuurkundige-, statistische- en machine learning-modellen.

Samenvattend: in dit proefschrift leveren wij een bijdrage aan twee centrale uitdagingen voor de inschatting van aardwetenschappelijke parameters: ontbrekende spatiëel-temporele informatie in de data, en meerdere passende oplossingen voor dezelfde observaties. Via vier specifieke contributies, beschreven in Hoofdstukken 3–6, hebben wij de kennis over deze uitdagingen verbreed, en de betrouwbaarheid van inschattingmethodes verbeterd. Hoewel de invloed van deze uitdagingen niet geheel verwijderd kan worden binnen het kader van een enkel proefschrift, vertegenwoordigt ons onderzoekswerk zowel concrete vooruitgang als uitvoerbare aanbevelingen voor verder onderzoek. Wij moedigen andere wetenschappelijk onderzoekers aan om verder te gaan in onze onderzoeksrichting, en om bij te dragen aan verder multidisciplinair onderzoek, met als doel om verbeteringen te realiseren in het inschatten van de status van de Aarde.