



Contribution ID: 33

Type: not specified

## Large-scale aerial photo processing for tree health monitoring with HTCondor

Friday 27 September 2019 09:25 (25 minutes)

```
\documentclass{article}
\usepackage{filecontents}
\usepackage{authblk}
\usepackage{natbib}
\usepackage{natbib}
\bibliographystyle{abbrvnat}
\setcitestyle{numbers,open={},close={},citesep={,}}
\begin{filecontents}{\jobname.bib}
@article{phillips2008a,
author = {Phillips, S.J. and Dudík, M.},
title = {Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation},
journal = {Ecography},
volume = {31},
number = {2},
year = {2008},
pages = {161–175},
url = {https://doi.org/10.1111/j.0906-7590.2008.5203.x},
language = {en}
}
@article{phillips2017a,
author = {Phillips, S.J. and Anderson, R.P. and Dudík, M. and Schapire, R.E. and Blair, M.E.},
title = {Opening the black box: an open-source release of Maxent},
journal = {Ecography},
volume = {40},
number = {7},
year = {2017},
pages = {887–893},
url = {http://doi.org/10.1111/ecog.03049},
language = {en}
}
@article{Soille,
author = {Soille, P. and Burger, A. and De Marchi, D. and Kempeneers, P. and Rodriguez, D. and Syrris, V. and Vasilev, V.},
title = {A versatile data-intensive computing platform for information retrieval from big geospatial data},
journal = {Future Gener. Comput. Syst},
volume = {81},
year = {2018},
pages = {30–40},
url = {https://doi.org/10.1016/j.future.2017.11.007},
language = {en}
}
@book{beck2019,
author = {Beck, P. S. A and Martínez-Sánchez, L. and Di Leo, M. and Chemin, Y. and Caudullo, G. and de la
```

```

Fuente, B. and Zarco-Tejada, P. J.},
title = {The Canopy Health Monitoring (CanHeMon) project},
institution = {Joint Research Centre (European Commission)},
publisher = {Publications Office of the European Union, Luxembourg},
year = {2019},
isbn = {978-92-79-99639-9},
doi = {10.2760/38697},
pages = {79},
language = {en},
}
@book{marti,
author = {Beck, P. S. A and Martínez-Sánchez, L. and Di Leo, M. and Chemin, Y. and Caudullo, G. and de la
Fuente, B. and Zarco-Tejada, P. J.},
title = {Remote Sensing in support of Plant Health Measures–Findings from the Canopy Health Monitoring},
institution = {Joint Research Centre (European Commission)},
publisher = {Publications Office of the European Union, Luxembourg},
year = {2019},
isbn = {978-92-76-02051-6},
doi = {10.2760/767468},
pages = {13},
language = {en},
}
@article{haralick1973a,
author = {Haralick, R.M. and Shanmugam, Kand Dinstein,I.},
title = {Textural Features for Image Classification},
journal = {IEEE Transactions on Systems, Man, and Cybernetics},
volume = {3},
year = {1973},
pages = {610–621},
url = {https://doi.org/10.1109/TSMC.1973.4309314},
number = {6},
language = {en}
}
@article{delafuente,
author = {de la Fuente, B. and Saura, S. and Beck, P. S. A.},
title = {Predicting the spread of an invasive tree pest: the pine wood nematode in Southern Europe},
journal = {Journal of Applied Ecology},
volume = {55},
year = {2018},
pages = {2374-2385},
doi = {10.1111/1365-2664.13177},
number = {5},
language = {en}
}
\end{filecontents}
\usepackage[utf8]{inputenc}

\begin{document}
\title{Large-scale aerial photo processing for tree health monitoring with HTCondor}

\author[1]{Martinez-Sanchez, Laura}
\author[1]{Rodriguez-Aseretto, Dario}
\author[1]{Soille, Pierre}
\author[1]{Beck, Pieter S. A.}
\affil[1]{European Commission, Joint Research Centre (JRC)}
\date{September 2019}
\maketitle
\begin{abstract}

```

The Canopy Health Monitoring (CanHeMon) project ran at the Joint Research Centre of the European Commission from mid-2015 to mid-2018 and was funded by DG SANTE. DG SANTE is responsible, among other things, for the European Union’s Plant Health legislation, which aims to put in place effective measures to protect the Union’s territory and its plants, as well as ensuring trade is safe and the impacts of climate change on the health of EU crops and forests are mitigated. For specific harmful organisms that threaten its crops and forests, the EU takes emergency control measures. The Pine wood nematode (*Bursaphelenchus*

xylophilus)) is such a quarantine pest. It can kill European coniferous tree species and is spreading through Portugal since the end of the 1990s.

As part of the EU emergency measures against the pine wood nematode (PWN) (\textit{Bursaphelenchus xylophilus}) Decision 2012/535/EU, Portugal should perform, outside and during the flight season of the PWN's vector, surveys of coniferous trees located in the 20 km wide buffer zone established along the Spanish border, with the aim to detect trees which are dead, in poor health or affected by fire or storm. These trees shall be felled and removed to avoid that they act as attractants for the longhorn beetle (*Monochamus* sp), the insect vector responsible for the spread of PWN \citep{delafuente}. The CanHeMon project tasked the Joint Research Centre with analysing a portion of the buffer zone, using remote sensing data, to support detection on the ground of declining pine trees. During the project, a 400 km<sup>2</sup> area was imaged twice, in autumn 2015 and autumn 2016, at 15 cm resolution from aircraft, and individual declining tree crowns were detected using a MaxEnt-based \citep{phillips2017a,phillips2008a}, iterative image analysis algorithm, the performance of which was gauged through visual photointerpretation. The scalability of the automated methods was then tested using an image mosaic of the entire buffer zone at 30 cm resolution.

We sought an image analysis platform that could efficiently handle and parallelise the computations on the large (terabyte) volumes of image data in this project. The JRC Earth Observation Data and Processing Platform (JEODPP) \citep{Soille}, which was developed in parallel with the CanHeMon project \citep{beck2019,marti}, increasingly met these needs over the course of the project. Being an inhouse service of the EC, it facilitates processing of the data for which the licensing does not permit public distribution. It is a versatile platform that brings the users to the data through web access and allows for large-scale batch processing of scientific workflows, remote desktop access for fast prototyping in legacy environments, and interactive data visualisation/analysis with JupiterLab.

The storage and processing nodes underlying the JEO DDP infrastructure consist of commodity hardware equipped with a stack of open source software. The storage service relies on the CERN EOS distributed file system which provides a disk-based, low latency storage service suitable for multipetabyte scale data. EOS is built on top of the XRootD protocol developed for high energy physics applications but also offers almost fully POSIX compliant access through a dedicated FUSE client called FUSEX that is suitable for other areas. As of summer 2019, the storage capacity of the EOS distributed file system of the JEO DDP amounts to 14 PiB corresponding to a net capacity of 7 PiB given that all data are replicated once to ensure their availability and decrease the likelihood of data loss in case of disk failure. For all other services, the JEO DDP relies on processing servers with a total of 2,200 cores distributed over 64 nodes. On average, 15 GB of RAM is available to each core. The batch processing service, called JEO-batch, is orchestrated with HTCondor. All applications running on the JEO DDP are deployed within Docker containers to ease the management of applications having conflicting requirements in terms of library versions. Docker images are created by combining and modifying standard images downloaded from repositories. For the canopy health monitoring application, we created a Debian image with all the libraries needed to run the code (mainly R and GDAL libraries).

Covering the entire PWN buffer zone with 4-band images of 30 cm resolution, stored in 8-bit, generates 2.4 TB of data. The associated texture layers \citep{haralick1973a} used in the analyses here added an additional 50 TB. The data were delivered and processed in 24,904 tiles measuring 1 km by 1 km. Processing a single tile in each iteration takes 40 to 55 minutes, with a memory usage of 5-7 GB on a regular CPU (with 2 to 8 cores). Processing the entire buffer zone on a single CPU would thus take more than a year. Assigning all of the 2,200 cores of the JEO DDP to batch processing service and submitting the job with HTCondor, the task would be completed in less than two hours. In practice, between 100 and 500 cores of the JEO DDP were used in the processing at any one time. The results of this processing were used to support management of the area on the ground and make recommendations on the use of remote sensing for large-area surveys in the context of plant health.

\end{abstract}

\bibliography{jobname}

\end{document}

## Desired slot length

15

## Speaker release

Yes

**Primary authors:** MARTINEZ SANCHEZ, Laura (JRC); RODRIGUEZ ASERETTO, Dario; Mr SOILLE, Pierre (European Commission, Joint Research Centre (JRC) Directorate I. Competences. Unit I.3 Text and Data Mining); Dr BECK, Pieter (European Commission, Joint Research Centre (JRC))

**Presenter:** MARTINEZ SANCHEZ, Laura (JRC)

**Session Classification:** Workshop presentations

**Track Classification:** HTCondor presentations and tutorials