An Automated Workflow for Satellite-based Monitoring of Field Flooding

Xinyi Wang, MASc., Wan-Chun Liao, Cody Carroll, Ph.D.
Data Institute, University of San Francisco

UNIVERSITY OF SAN FRANCISCO

DI DATA INSTITUTE · UNIVERSITY OF SAN FRANCISCO

Background and Problem Overview

Wetlands and the BirdReturns Program

Figure 1: Wetland Status Map for California Central Valley
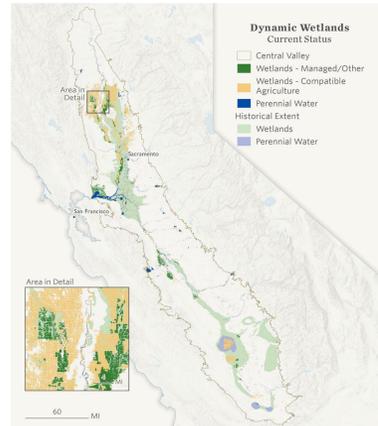
Every year, millions of ducks and shorebirds migrate along the Pacific Flyway and travel through California.[1] Many of these birds need wetlands for food and rest to support their journey.

However, over 90% of Central Valley's historical wetlands (light green and blue in Fig. 1) have been drained and developed.[2]

Fortunately, some of these wetlands have been converted to crops like rice (yellow in Fig. 1) that can support migratory wetland birds if managed appropriately.

The Nature Conservancy (TNC) and partners launched the BirdReturns program in 2014 to pay farmers to flood their fields to support migratory wetland birds. Simultaneously, BirdReturns delivers multiple benefits (see Ecological and Social Impact) for farmers, wetland managers, and communities across the Valley.[3]

The program created over 50,000 acres of flooded wetland habits in 2021-2022. The program aims to scale up and deliver an additional 100,000 acres and serve them as a rapid response program for maintaining habitat during drought. [3]

Read More: https://birdreturns.org/

Problem Definition

As the BirdReturns initiative scales up, new technology is needed to cost-effectively monitor the enrolled fields and ensure they are flooded. TNC has been experimenting with using free satellite images to estimate the extent and duration of flooding on the enrolled fields.

However, the established workflow involves significant manual work to run the Google Colab code and share the report. It is not adaptive to new fields or programs.

Our contribution is an automated data pipeline built to:

- ingest the satellite data,
- generate flooding extent estimates,
- construct visualizations of flooding trends over space and time, and
- send weekly reports to the program managers.

An Automated WorkFlow Approach

Our automated workflow extracts satellite images and filters by the fields' shapefiles and dates. Then flooding estimates are calculated using NDWI and summarized weekly. An interactive web dashboard with summary statistics and visualizations is then created and shared with stakeholders automatically by email.
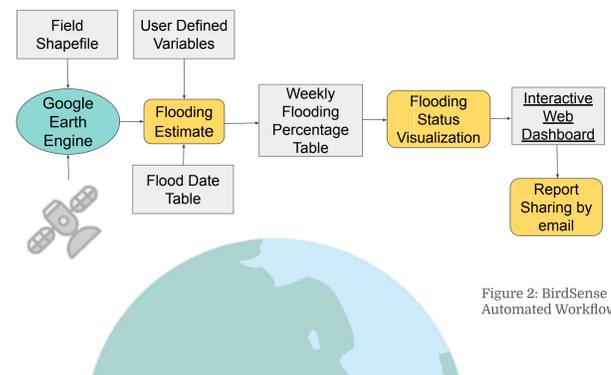

Figure 2: BirdSense Automated Workflow

Methods

Data Sources

Copernicus Sentinel 2 Satellite Images
The SENTINEL-2 mission is a European Earth observation program using a constellation of two satellites for optical imaging of land and coastal regions. It provides high-resolution images (10, 20 & 60 m) with 13-band data and high revisit frequency (2-3 days at mid-latitudes), which are widely used for land cover classification and changes monitoring. [4]
Sentinel 2 data can be accessed from Google Earth Engine (GEE) API through Google Cloud Service Account.

Google Drive for User-Provided Data
Users can save additional information, such as the flooding periods, in Google Drive as Excel, CSV files, etc.
The Google Drive Python API is used to download files from GDrive.

Estimating Flooding Extent

There are several approaches to estimating flooding extent using remote sensing (satellite data): threshold segmentation with Normalized Difference Water Index (NDWI) and image classification with machine learning algorithms. [5]
The NDWI approach proposed by Gao in 1996 [6] (equation below) was implemented using the Near-Infrared (NIR) and Short Wave Infrared (SWIR) channels.

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR}$$

where, NIR: Band 8 - 0.83 μm wavelength

SWIR: Band 11 - 1.61 μm wavelength

By choosing a threshold value, the water body can be extracted.

Building an Automated Workflow

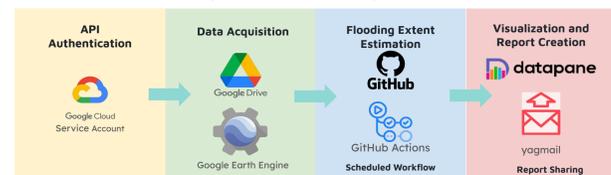The tools and techniques used for this project include:

Figure 3: Techniques Integrated into the Automated Workflow

Interactive Web Dashboard

A dashboard is created using DataPane and allows us to transform a Jupyter Notebook or Python script into an interactive web app.
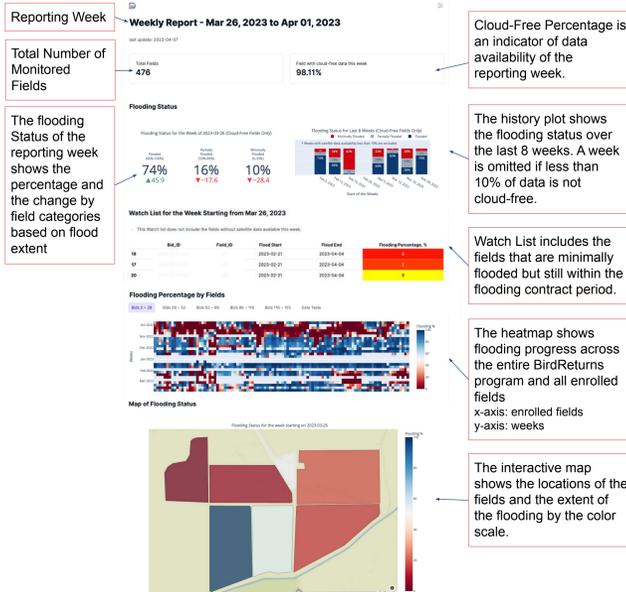

Figure 4: An Example of Interactive DataPane Dashboard.
The GitHub repository for this dashboard is publicly available at https://github.com/tnc-ca-geo/birdsense

Ecological and Social Impact

California's Central Valley is one of the largest agroecosystems in the world and serves critical purposes for California like addressing water scarcity, maintaining the freshwater ecosystem, and protecting migratory species.[7]
The BirdReturns program provides many ecological and social benefits to the Central Valley, including:

- Creating habitat and food sources for migratory birds.
  - Supporting over 1 million migratory birds across 50+ species [8]
- Improving shorebird density, richness, and diversity.[1]
- Reducing water management cost:
  - The renting land cost is only 0.5%-1.5% of the purchasing and maintaining cost to create migratory bird habitats. [7]
- Benefits to the local community:
  - Recharging groundwater and improving water security.[7]
- Benefits to farmers:
  - Preparing farmlands by flooding irrigated crops, flushing salts, and decomposing stubble. [7]
  - Generated over $2 million in direct payments to farmers. [3]

Currently, the automated workflow is being used as a complement to the expensive and time-consuming on-the-ground surveying. Our goal is to eventually have it replace the in-person monitoring to reduce costs and improve consistency and efficiency in field assessment.

Discussion

Pros of the Automated BirdSense Workflow

- Reduces workload: no human intervention is required
- Lowers the cost of monitoring field flooding:
  - Free satellite data instead of traditional ground-based survey
  - GitHub Actions allows 2,000 minutes of run time per month
  - GEE, Google Drive, and DataPane are free
- Easy to set up: no need to set up the environment by yourself
- Local machine agnostic: run on GitHub-hosted runners (Linux)
- Flexible: adaptable to new programs and functions
- Secure: sensitive information (tokens) saved in repo Secrets

Comparison with Other Automation Tools

We also considered using alternative automated data pipeline tools, such as Airflow, Stitch, FiveTran, etc.[9] Apache Airflow is a popular open-source platform for creating, scheduling, and monitoring workflows. We established a workflow using Airflow as an alternative. We compared Airflow and GitHub Actions for the automation purpose and summarized our findings below:

Features	GitHub Actions	Apache Airflow
Use Cases	CI/CD platform Simple data pipelines	Complex data pipelines with many tasks and dependencies
Complexity	Relatively Low. Easy to set up	High. Steep learning curve
Customization	Limited options	Highly customizable and extendable
Scalability	Relatively low. May not handle large-scale data pipelines	High. Scalable solution for large data pipelines and growing data volume.
Integration	Integrates well with GitHub repositories	Widely integrated with data stores, other services, machine learning frameworks, etc.
Infrastructure Requirement	Local machine agnostic. Run on GitHub-hosted runners (Linux)	Local or virtual machine
Reproducibility	Easy to share and reproduce	Environment prerequisite for reproduce
Cost	Free usage with a run-time limit	Free and open source
Security	Allow to save tokens in repo secrete.	Require third-party services to store and access tokens and passwords

Table 1: Comparison of GitHub Action and Apache Airflow

GitHub Actions provides a centralized, streamlined, and lightweight solution for simple automation tasks. It can be easily integrated into existing workflows and set up with minimal effort. On the other hand, Airflow is a powerful tool for managing complex workflows. It requires significant infrastructure and resources to set up and maintain.

Future Applications of the Workflow

Our next step is to apply this automatic workflow to new BirdReturns programs and other TNC projects requiring automation. The workflow also provides an easy and cost-effective way to build a data pipeline for tasks like data acquisition through API, satellite image analysis, dashboard development, or report sharing. This approach is especially beneficial for small-scale projects and start-up companies, for which efficiency and costs are the primary concerns.

Acknowledgements

References

1. Golet, Gregory H., et al. "Using ricelands to provide temporary shorebird habitat during migration." Ecological Applications 28.2 (2018): 409-426.
2. Reiter, Matthew E., et al. "Local and landscape habitat associations of shorebirds in wetlands of the Sacramento Valley of California." Journal of Fish and Wildlife Management 6.1 (2015): 29-43.
3. The Nature Conservancy, "BirdReturns - Creating flexible bird habitat in California's Central Valley," 2023. [Online]. Available: https://birdreturns.org/.
4. The European Space Agency, "SENTINEL-2 MISSION GUIDE," [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2.
5. Li, J., Ma, R., Cao, Z., Xue, K., Xiong, J., Hu, M., & Feng, X. (2022). Satellite Detection of Surface Water Extent: A Review of Methodology. Water, 14(7), 1148. https://doi.org/10.3390/w14071148
6. Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3), 257-266. https://doi.org/10.1016/S0034-4257(96)00067-3
7. Rohde, Melissa M., Mark Reynolds, and Jeanette Howard. "Dynamic multibenefit solutions for global water challenges." Conservation Science and Practice 2.1 (2020): e144.
8. Sesser, Kristin A., et al. "Waterbird response to variable-timing of drawdown in rice fields after winter-flooding." Plos one 13.10 (2018): e0204800.
9. TrustRadius, "Data Pipeline Tools," [Online]. Available: https://www.trustradius.com/data-pipeline