Agenda

1. Who we are
2. Predicting flood with machine learning
3. Building a resilient machine learning product
Better decisions faster and cheaper

Collect data from across your infrastructure to move from reactive to proactive asset management, optimising performance and cutting costs.

Read more

Turn data into information

Moata is a cloud-based service which provides stakeholders with access to real-time asset information. Its intuitive interface makes data easy to understand and action.

Read more
Background
Flooding is an urgent problem

Cost to the NZ insurance industry due to flooding ($ million)

(Insurance Council of New Zealand, 2020)
What if we can reliably predict flooding in advance?
Background
The challenges

1. Unavailability of historical data
2. Immediate effect of rainfall
3. Needs for a scalable prediction framework
4. A complete system that performs data gathering, processing, computation and alarm
Background

Complete solutions that consider all practical aspects are rather limited

Publicly available works either

- **Provide predictions in 15-minute to 1-hour resolution or multiple hours to daily resolution**
  → may not be granular enough to capture urban flow spikes;

- **Provide predictions in limited forward-looking window, from 0-minute to 1-hour ahead**
  → allows little time to carry out counter measures

- **Relies mainly on the delayed effect from upstream flows**
  → not applicable for locations with limited or no upstream, or has immediate reaction to rainfall

- **Utilise computationally expensive hydraulic models**
  → not scalable for large scale deployment
### Our contribution

<table>
<thead>
<tr>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>30s to calibrate and sub-second prediction</td>
</tr>
<tr>
<td>Always learn</td>
</tr>
<tr>
<td>2cm accuracy</td>
</tr>
<tr>
<td>Integrated</td>
</tr>
<tr>
<td>Scalable deployment to 1000's of sites, simulations of scenarios</td>
</tr>
<tr>
<td>Stay relevant</td>
</tr>
<tr>
<td>Efficient planning</td>
</tr>
<tr>
<td>User friendly</td>
</tr>
</tbody>
</table>
Methodology

Data gathering, storage, retrieval
Methodology

Data preprocessing

1. Aggregate
   Aggregate data to 5 minute frequency

2. Clean
   Removing abnormal periods using statistical tests and domain knowledge

3. Cross correlation
   Determine and account for the delayed effect of rainfall
Methodology

Rainfall prediction using Short Term Ensemble Precipitation

1. Probabilistic nowcasting system

2. Generate estimates of possible future rainfall

3. Very accurate in the 0-6 hour range
Methodology

Flow and level prediction
## Results

### Summary

<table>
<thead>
<tr>
<th>No.</th>
<th>Site</th>
<th>Mean Absolute Error (mm)</th>
<th>Mean Absolute Percentage Error (%)</th>
<th>Model training time (seconds)</th>
<th>Time to output prediction (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Western Springs</td>
<td>27.13</td>
<td>5.24</td>
<td>28</td>
<td>0.12</td>
</tr>
<tr>
<td>2</td>
<td>Motions Road Weir</td>
<td>25.64</td>
<td>4.34</td>
<td>41</td>
<td>0.11</td>
</tr>
<tr>
<td>3</td>
<td>New Market</td>
<td>18.22</td>
<td>5.17</td>
<td>34</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td><strong>Average</strong></td>
<td><strong>23.66</strong></td>
<td><strong>4.75</strong></td>
<td><strong>37.5</strong></td>
<td><strong>0.11</strong></td>
</tr>
</tbody>
</table>
Results

By events

Stream Flow ARI

Stream depth
2 mins maximums
- MIN: 0.18m
- MAX: 1.32m
- MEAN: 0.44m

Stream depth - 6H prediction
2 mins maximums
- MIN: 0.18m
- MAX: 1.48m
- MEAN: 0.48m

ACC - Rain - Cox's Bay Park Rainfall

Rainfall
1 hr totals
- MIN: 0.50mm
- MAX: 13.00mm
- TOTAL: 8750mm

Ty108 Peak ARI
- 12 hrs: 2.37y

3 April 2017, 09:22 to 7 April 2017, 07:17
Results

By events

Stream Flow ARI

Stream depth
2 mins maximums
- MIN: 0.22m
- MAX: 1.15m
- MEAN: 0.30m

Stream depth - 6H prediction
2 mins maximums
- MIN: 0.22m
- MAX: 1.28m
- MEAN: 0.35m

ACC - Rain - Cox's Bay Park Rainfall

Rainfall
1 hr totals
- MIN: 0.50mm
- MAX: 12.00mm
- TOTAL: 29.00mm

Rainfall (Accumulated)

4 September 2019, 11:49 to 7 September 2019, 23:35
Results

By events
Conclusion

This framework offers a significant tool for improving flood protection by enabling a rapid preventive approach which can be rolled out in scale.

01 **Usability**
Provides granular (5 minutes) prediction for a 6-hour forward looking window with high accuracy, which allows reasonable time to react to overflow events.

02 **Scalability**
Each ML model takes approximately 30 seconds to train and less than a second to output prediction for the next 6 hours, which allows scalability to thousands of sites with little effort.

03 **Account for uncertainty**
Predict flow in different simulated rainfall scenarios and utilize ensemble method in ML to consolidate them into a single ‘best’ prediction.
MLOps

Building a resilient machine learning product
What is MLOps?

Framework that integrates Machine Learning with DevOps.

The concept is integral when putting Machine Learning models in production.

Generally carried out after the model has been trained and validated.
How did MLOps come to be?

The term was coined in a 2018 presentation from Google executives.

It is estimated by Deloitte that the number of Machine Learning (ML) implementations in organisations quadrupled between 2017-2020.

In 2017 McKinsey released a study estimating that 88% of ML processes do not end up in production.

Companies that did manage to put ML models in production saw 3-15% profit margin increases.
Why use MLOps?

- CI/CD allows for automatic, and unit tested deployment of models
- Data drift triggers can be used to retrain model and manage performance
- Highly adaptable once infrastructure has been set up
CI/CD

Continuous Integration
We want all parts of what goes into training and deploying a model to be in the same place and with easy access.

Commit all your code to a single repository!

Implement processes once a change is detected:
- Code Quality tests
- Unit tests
- Test coverage
CI/CD

Continuous Deployment

Deployment is hard! One of the easiest things to do is automate it.

Once changes to your repository have been made and tests have passed, your model can be deployed automatically to a development or production environment for usage.
CI/CD

Traditionally

CI/CD

Mott MacDonald | Smart Infrastructure
Data Drift

Defined as a change between live-production data and a training dataset.

Production data can diverge due to different reasons.

- **Concept drift**: Change in the relationship between the model input and output.
- **Target drift**: Change in the model’s output or label distribution.
- **Feature drift**: Change in the input data distribution.

All these scenarios should trigger a retraining of the model. With MLOps this process is automated.
Our Framework
Thank you