Confronting physical models with noisy data

Using satellite imagery to calibrate Bayesian wetland models

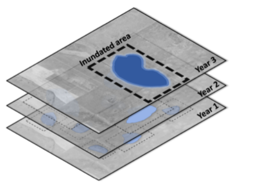
Christopher Krapu[1] christopher.krapu@duke.edu
Mark Borsuk[1] mark.borsuk@duke.edu
Mukesh Kumar[1,2] mkumar4@ua.edu

[1]Department of Civil & Environmental Engineering, Duke University [2]Department of Civil, Construction & Environmental Engineering, University of Alabama

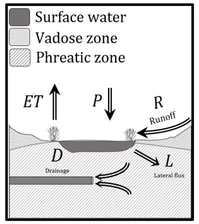
Background

Motivation

- Humans live amongst and make use of **millions of lakes and wetlands** across the planet with a high concentration of them located in the **Prairie Pothole Region** in central North America.
- They are **too numerous to measure with in-situ instrumentation** but their inundated surface area can be observed with satellite imagery programs such as Landsat and Sentinel.
- Effective management of inland water bodies is greatly aided by **models which can be used for forecasting and monitoring.**
- **Calibrating conventional hydrology models with remote sensing is difficult** as this data's spatial and temporal resolution is often much coarser than with in-situ gauges and often exhibits systematic bias.



The hydrologic state of wetland and lakes can be partially observed via satellite imagery of their surficial extent.



In the absence of connections to a drainage network, prairie wetlands' water budgets are dominated by runoff from snowmelt and evapotranspiration.

Methods

Data

- To develop and assess our model, we required **overlapping in-situ and remote sensing measurements of water levels**
- We studied three wetlands (P01, P06 and P07) from the **Cottonwood Lake Study Area**[1] in North Dakota, USA. This is a USGS field site with long term water level records.
- We also used the **Global Surface Water**[2] **data product** which provides estimates of water presence/absence across Earth's surface at monthly intervals and is derived from Landsat images over the years 1984 – 2015.
- We used **gridMET**[3] **estimates of precipitation, potential evapotranspiration and temperature** for model forcing.

The wetlands of CLSA are exemplars of the Prairie Potholes as they exhibit substantial variation in inundated extent due to precipitation and evapotranspiration.

Model

- Our model is based on **a water balance equation** which expresses the change in ponded water volume v_t as the sum of a positive term due to runoff and direct precipitation with a loss term dependent on the area and perimeter of the ponded water area for the *i*-th wetland.
- The loss function l_t is parameterized by **power-law functions of the volume**. This power-law representation is motivated by geometric considerations regarding the bathymetry of a wetland and the topography of its surrounding slope.
- We account for consistent underestimation of true water surface area by assuming a **skew-normal distribution for the measurement error** $\boldsymbol{\epsilon_t}$.

$$g_t = p_t a_t + c(rp_t + s_t)$$

Gain in volume — Direct precipitation onto surface area — Runoff and snowmelt over the catchment

$$l_t = e_t(a_t + b_t)$$

Loss in volume — Perimetric loss due to lateral groundwater flow and areal loss to ET

$$a_t = \alpha_1 v_t^{\beta_1}$$

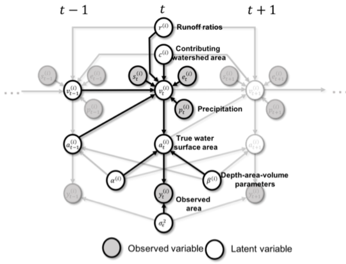
Surface area — Power-law function of volume

$$b_t = \alpha_2 v_t^{\beta_2}$$

Perimetric loss — Power-law function of volume

$$y_t = a_t - \epsilon_t$$

Observed surface area — True area — Skew error process



Parameter estimation and inference

- Inference for noisy nonlinear dynamical systems is known to be **very difficult due to correlations between parameters in the posterior** $p(\theta|\mathcal{D})$.
- Modern varieties of Markov chain Monte Carlo can efficiently draw samples from correlated posterior distributions **using gradients of the log-posterior** $\boldsymbol{\nabla_\theta \ln p(\theta|D)}$.
- Since these gradients are taken across many timesteps of the model, **we use automatic differentiation as implemented in Theano**[4] to perform these calculations which would be impossible to write in closed form.
- After writing our wetland hydrology model in Theano, **we used the No-U-Turn sampler**[5,6] **to draw 4000 posterior samples** for the parameters of each wetland

Parameter	Site	Posterior mean	95% credible interval		N_{eff}	$\hat{R}$
			Lower bound (2.5%)	Upper bound (97.5%)		
α_1	P01	0.31	0.15	0.46	555	1.002
	P06	0.27	0.14	0.41	850	1.001
	P07	0.29	0.14	0.42	546	1.000
β_1	P01	0.42	0.33	0.50	551	1.001
	P06	0.39	0.30	0.49	834	1.001
	P07	0.39	0.29	0.49	562	1.000
α_2	P01	0.52	0.04	1.01	267	1.003
	P06	0.24	0.00	0.58	114	1.010
	P07	0.23	0.00	0.53	453	1.003
β_2	P01	0.02	0.00	0.06	825	1.000
	P06	0.04	0.00	0.12	972	1.000
	P07	0.03	0.00	0.11	782	1.001
c	P01	5.48	4.32	7.05	236	1.001
	P06	4.67	2.56	5.94	231	1.001
	P07	4.84	3.06	6.54	275	1.000
r	P01	0.87	0.68	0.99	410	1.001
	P06	0.48	0.25	0.84	205	1.001
	P07	0.48	0.26	0.80	281	1.001
v_0	P01	0.25	0.03	0.70	1592	1.000
	P06	0.26	0.03	0.72	329	1.003
	P07	0.25	0.05	0.65	1603	1.001

Summary of samples drawn from the posterior distribution

Solution

- A **statistical modeling framework** allows us to account for **uncertainty** and correct **systematic biases**.
- Standard statistical techniques such as time series models and regressions **do not make use of our extensive prior knowledge** of the mechanisms governing inflow and outflow of water.
- To realize benefits from both physical and empirical modeling techniques, **we embedded a hydrological model within a Bayesian statistical framework**.
- Bayesian models require assumption of **prior distributions** over all parameters in tandem with identification of a **likelihood function** mapping parameters to observed data.

Bayes' Theorem

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int_\theta p(\mathcal{D}|\theta)p(\theta)d\theta}$$

Model parameters θ Effective catchment area, Runoff ratio

Likelihood function Wetland hydrology model

Prior assumptions

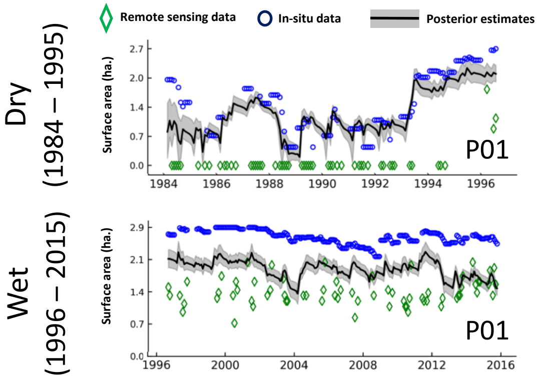
Posterior distribution Describes probability different parameter values conditional on observations

Observed data $\mathcal{D}$ Monthly inundated surface area

Marginal probability Computationally intractable and not available in closed form

Results

- We assessed our model by determining whether it could accurately **estimate the true wetland water area and volume** with only **noisy observations**.
- While the model estimates were highly accurate during a regional dry period (1984 – 1995), they were very inaccurate during periods of increased precipitation (1996 – 2015).



Conclusions

- We developed a wetland hydrological model that can be calibrated with **only satellite imagery**
- Accurate characterization of wetland water levels **during dry periods** was observed despite substantial observation bias
- Fluxes from the groundwater table into the ponded area during extended wet periods **are not well represented**

Acknowledgements

This work was supported by the Institute for International Applied Systems Analysis (IIASA) via the Young Scientist Summer Program and by a NASA Earth and Space Sciences Fellowship.

References

1 Winter, Thomas C., and Donald O. Rosenberry. "Hydrology of Prairie Pothole Wetlands during Drought and Deluge: A 17 Year Study of the Cottonwood Lake Wetland Complex in North Dakota in the Perspective of Longer Term Measured and Proxy Hydrological Records." *Climate Change* 40 (1998): 189–209.
2 Pekel, Jean-François, Andrew Cottam, Noel Gorelick, and Alan S. Belward. "High-Resolution Mapping of Global Surface Water and Its Long-Term Changes." *Nature* 540, no. 7633 (December 7, 2016): 418–22. https://doi.org/10.1038/nature20584.
3 Abatzoglou J. T., Development of gridded surface meteorological data for ecological applications and modelling, International Journal of Climatology. (2012) doi: https://doi.org/10.1002/joc.3413
4 Bergstra, James, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. "Theano: A CPU and GPU Math Compiler in Python," 2010, 7.
5 Hoffman, Matthew D., and Andrew Gelman. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.* 15, no. 1 (January 2014): 1593–1623.
6 Salvatier, John, Thomas V. Wiecki, and Christopher Fonnesbeck. "Probabilistic Programming in Python Using PyMC3." *PeerJ Computer Science* 2 (April 6, 2016): e55. https://doi.org/10.7717/peerj-cs.55.