9th International Workshop on Climate Informatics
October 02-04, 2019
Hosted by École Normale Supérieure, Paris, France

EMULATING NUMERIC HYDROCLIMATE MODELS WITH PHYSICS-INFORMED CGANS

Ashray Manepalli[1], Adrian Albert[1,2,*], Alan Rhoades[2], Daniel Feldman[2], Prabhat[2]

Abstract—**Process-based numerical simulation, including for climate modeling applications, is compute- and resource-intensive, requiring extensive customization and hand-engineering for encoding governing equations and other domain knowledge. On the other hand, modern deep learning employs a much simplified and efficient computational workflow, and has been showing impressive results across myriad applications in computational sciences. In this work, we investigate the potential of deep generative learning models, specifically conditional Generative Adversarial Networks (cGANs), to simulate the output of a physics-based model of the spatial distribution of the water content of mountain snowpack, or snow water equivalent (SWE). We show preliminary results indicating that the cGANs model is able to learn mappings between meteorological forcing (e.g., minimum and maximum temperature, wind speed, net radiation, and precipitation) and SWE output. Moreover, informing the model with simple domain-inspired physical constraints results in higher model accuracy, and lower training time. Thus Physics-Informed cGANs provide a means for fast and accurate SWE modeling that can have significant impact in a variety of applications (e.g., hydropower forecasting, agriculture, and water supply management).**

I. MOTIVATION

In many climate modeling applications, direct observation on large scales of the environmental variables of interest is challenging, requiring instead the use of computationally-expensive numerical simulation models [1]. Such numerical weather and climate models are based on a series of coupled partial differential equations (PDEs) that aim to represent the dynamics, thermodynamics, radiative, and mass-flux processes within the major components of the Earth system including the atmosphere, cryosphere, land-surface, and ocean. These PDEs are often representative of the forefront of scientific understanding - utilizing fundamental physics, hydrology, and climatology theory - but are computationally-expensive to solve, requiring the use of high-performance computing (HPC) environments and highly-specialized expertise to set up and operate [2].

[1] terrafuse, inc. [2] Berkeley National Lab
*Correspondence to: toni@terrafuse.ai

In addition, such process-based models of realistic systems often employ parametrizations to resolve sub-grid processes that are generally poorly understood, making it hard to decipher model sensitivity and bias, especially when all components of the Earth system are coupled [2]. One such variable is the aforementioned SWE, requiring the use of a chain of expensive numerical models for simulation. Empirical SWE data is typically highly sparse or even completely unavailable due to its difficulty in acquisition from mountainous regions, as well as the high expense associated with maintaining measurements at adequete temporal and spatial resolutions [3]. Moreover, SWE has many important use cases across sectors of high societal impact, e.g., water supply, hydropower, and agriculture [4].

Challenges in snowpack modeling. We focus on two current shortcomings of process-based models. First, the forcing uncertainty in key meteorological variables, including precipitation amount and phase, air temperature, and humidity, is shown to be comparable to or larger than snowpack model structural uncertainty [5]. Second, snow models heavily rely on temperature dependent thresholds to determine the phasing of incident precipitation and the magnitude and duration of the cold content of snow, or the interplay between snow density, depth, and temperature prior to melt. Therefore, a key outstanding need in the community would be to test how biases in precipitation intensity, duration and frequency and phase drive divergence in the snowpack accumulation season and how biases in surface energy and mass flux drive early spring melt [6].

II. METHODS, DATA, AND MODELS

Data. For all experiments presented here we have used a reanalysis dataset developed by Livneh[8] (L15) for the California Sierra Nevada mountain range. The L15 data was originally obtained by combining hydrologic simulation runs of the Variable Infiltration Capacity (VIC) model bounded by spatially interpolated in-situ meteorological station measurements. This dataset contains meteorological data and simulated SWE, used to train, assess, and constrain cGAN model. All data

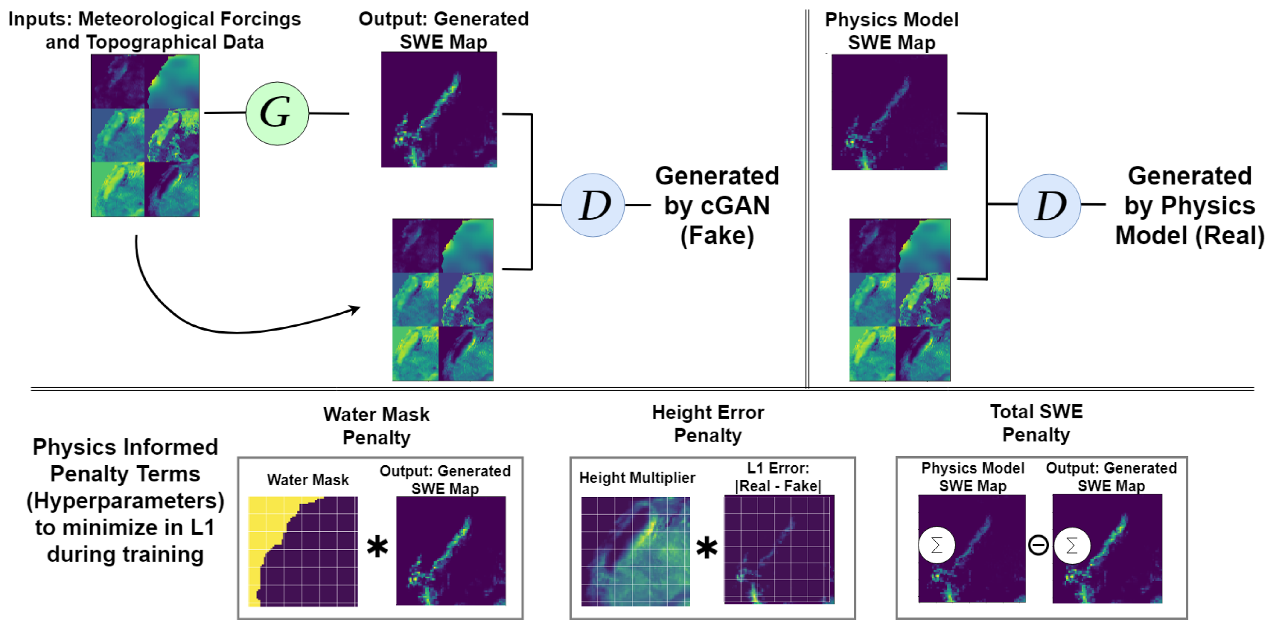

Fig. 1. Architecture and diagram of the conditional GAN used, a heavily modified variant of Pix2Pix [7]

channels are resized and normalized from 321x321 (4km grid size) to 64x64 (17km grid size).

Physics-informed conditional GANs. We formulate our emulation problem as an image-to-image translation task. The goal is to transform an image from domain X, gridded meteorological variables, to domain Y, SWE grids. The pipeline of training a GAN emulator of SWE is illustrated in Figure 1. In our setting, training samples from the two domains X and Y are assumed paired, $\{(x^i, y^i)\}_{i=1}^{N}$ as in [7]. Here we denote by x samples from domain X and by y samples from domain Y.

We have incorporated certain domain knowledge into our model via additional penalty terms into the optimization loss function, as follows:

- Areas of higher elevation typically have larger amounts of snow (and therefore SWE), and we add penalties to large errors in such areas accordingly;
- As a significant portion of the data we study covers water areas such as the Pacific Ocean, where no snowpack can exist, we penalize the model harshly for placing SWE values in these areas;
- We penalize the difference in total SWE between cGAN solutions and physics model output, to ensure that total stored water mass is properly estimated.

Training details. As in [9], we modified the standard GAN training scheme by first training the generator purely on L_1 loss term to estimate the conditional mean (for the first 5 epochs), and later adding an adversarial loss term to teach the generator finer details. We have also observed that this slight modification enables faster convergence to better solutions (with lower overall loss values). All deep learning training and inference was performed on a single consumer Graphical Processing Unit (GPU), the NVIDIA GTX 1080ti.

III. EVALUATION

Having trained the cGAN model as described above on a training set of 8 years of data (input/output pairs as described above at daily resolution), we have first tested its performance on a holdout sample of two years of data. This is a standard regression setting, for which we compute typical performance metrics. Even in this much simplified setting where we don't explicitly model time, the model achieves a mean absolute percentage error (MAPE) of 9.54%, indicating that it has learned a reasonably accurate mapping from meteorological and topographical data to simulated SWE.

In Figure 3 we show a comparison between cGAN and physics-based model output over 2-week periods at the end (June/August), start (November), and peak (April), respectively, of the SWE season (left, middle, and right panels in the figure, respectively). These are key periods of interest to mountain snowpack researchers and water resource managers, as they are check-in points in the lifecycle of mountain snowpack dynamics. We show the histograms of normalized pixel values of cGAN output (green line) and physics-based model output (black line). Note that the cGAN model

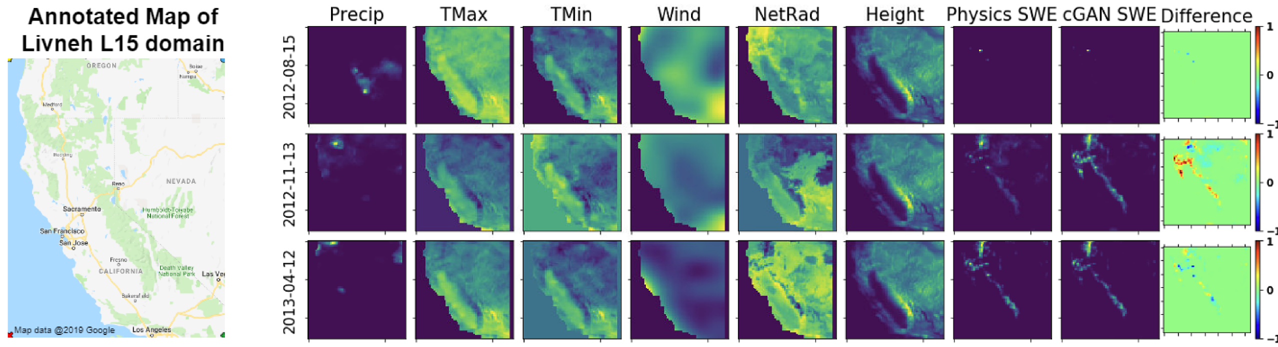

Fig. 2. Three samples (per row) of model inputs (meteorological forcings) on the first 6 columns, physics model output (column 7). cGAN output (column 8) and difference between physics model and cGAN (column 9). Rotated to match Sierra Nevada range (left).

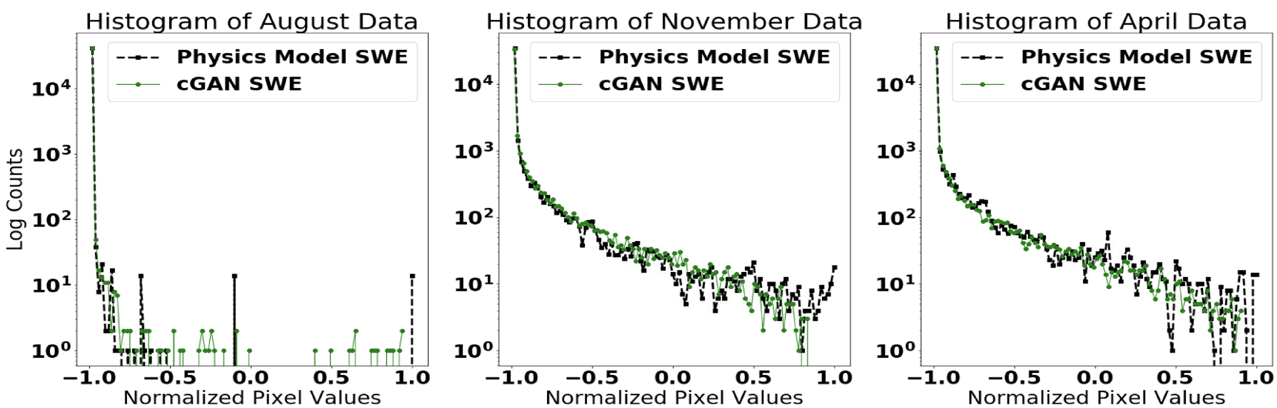

Fig. 3. Histograms of normalized pixel values comparing cGAN (green) and physics-model (black) across key snowpack seasons.

is able to accurately recover the distributions of values of the physics-based model.

Next, we calculate and plot the power spectral density function (PSD) for both the cGAN and physics-based model output. This metric incorporates information across all spatial frequency scales, and is defined as:

$$PSD = 10 \log_{10} |\mathcal{F}(\rho(SWE, SWE))|^2, \qquad (1)$$

where $\mathcal{F}$ denotes the Fourier transform and ρ the correlation coefficient, defined as usual. In Figure 4 we show the PSD profile comparisons for the key SWE seasons. Here too we observe strong performance - the spectral properties of the outputs the cGAN and the physics model are very similar, indicating that the cGAN performs well not just at 'memorizing' averages, but is also capable of recreating high frequency details.

Lastly, we have validated our hypothesis that inference time with a trained cGAN is extremely fast, taking less than 10 seconds to generate over 1000 simulated SWE grids on a GTX 1080ti, a consumer GPU. This suggests a speedup factor of around 1000x compared to just the raw runtime of a VIC model used to generate the SWE grids, which by our estimates takes $\sim$ 100 core-hours to simulate 100 years of SWE output. This speedup will allow for a much faster iteration to analyze how changes in meteorological forcing lead to changes in SWE, a topic of future research.

IV. INCORPORATING PHYSICS: ABLATION STUDY

To understand the effect of the physics-informed constraints and other inputs to our model, we performed an ablation study. We found that the inclusion of one input channel in particular, Net Radiation (NetRad) - a measure of the difference between incoming and outgoing atmospheric radiative energy - increased performance over all measured metrics. This matches up well intuitions from atmospheric science, where satellite radiometer data is used for the estimation of snowpack variables. We also find that our physics informed penalties improve performance on all metrics and convergence rates. This falls in line with physical intuition - penalizing physically incoherent solutions results in better performance.

We found that including 'physics penalties' improved performance across metrics. The inclusion of a penalty for the GAN assigning snow on known water areas

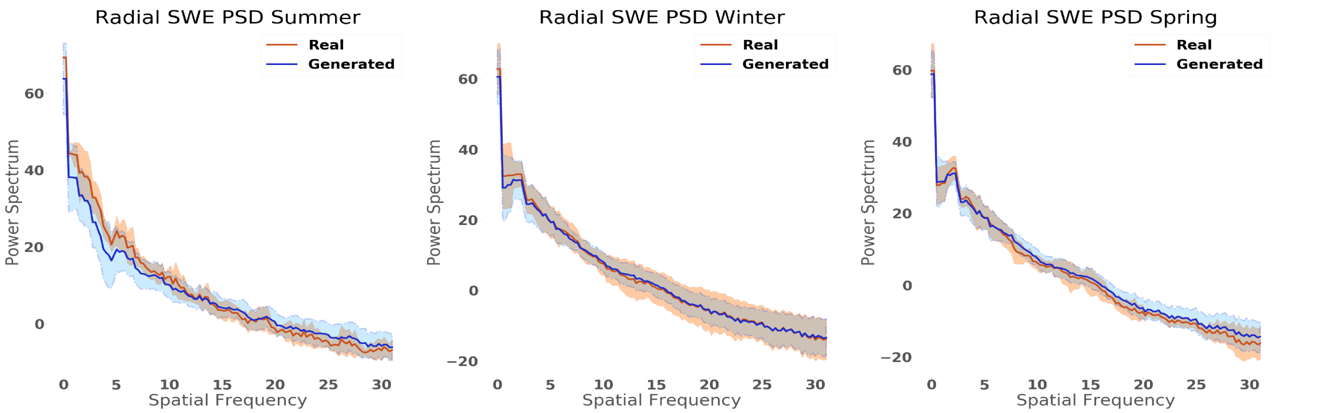

Fig. 4. Power Spectral Density of cGAN and Physics Model over different hydrologic seasons: respectively the end (Summer: July/August), start (Fall/Winter: November/December), and peak (Spring: April/May) of SWE season.

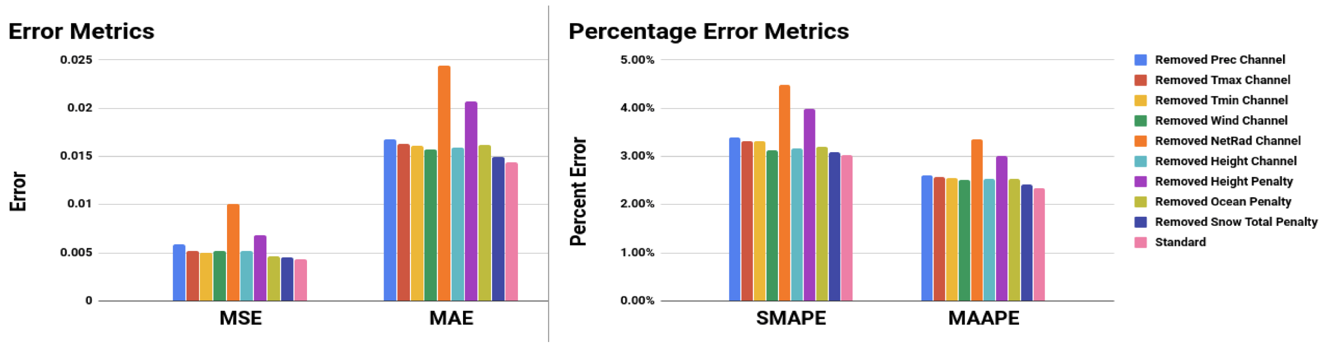

Fig. 5. Error metrics computed during ablation study. Each bar label represents the penalty or channel that was removed for the corresponding experiment. We observe that the inclusion of the Net Radiation channel and Height penalty both improve model quality, as the models excluding them resulted in significant increases in error.

greatly improved convergence during training. The inclusion of the height-based penalty of SWE errors at high altitudes made the model far stronger at generating sparse gridded SWE outputs and at recreating the tails of distributions as seen in Figure 3. The penalty on total SWE error did result in a slight increase in mean error, but forced the model to generate solutions with stronger physical coherence than as indicated by MAPE alone.

Figure 5 contains representative metrics logged for each parameter combination in the ablation study. Metrics computed are Mean Squared Error (MSE), Mean Absolute Error (MAE), Symmetric Mean Absolute Percent Error (SMAPE), and Mean Arctangent Absolute Percent Error (MAAPE). The traditional Mean Absolute Percent Error (MAPE) was also computed, but we found it not useful as a metric for model performance in SWE due to the frequently sparse maps - 'actual' data points are frequently zero, resulting in large error amplification. We recommend SMAPE and MAAPE as alternative metrics to correct this very problem.

V. FUTURE WORK.

The utility of deep learning for simulation of complex spatio-temporal systems such as SWE lies not just in the ability of models like the cGAN to quickly and efficiently produce accurate SWE predictions, but also in the ease of augmenting models and training schemes to better incorporate domain specific intuition. Such knowledge could be incorporated as "soft" constraints on training as done here, or as "hard" constraints directly incorporated into the design of the deep learning architectures - in both cases encouraging the model to generate physically coherent solutions.

In addition, our work so far has modeled SWE as an $i.i.d.$ process, focusing only on spatial correlations without any explicit temporal modeling. Since SWE is fundamentally a complex spatio-temporal process, new architectures must be designed explicitly around coherently modeling both its spatial and the temporal aspects. Furthermore, the current cGAN based approach of generating SWE maps is purely deterministic and does a poor job of modeling the uncertainty and stochasticity associated with hydrometeorological variables. Several

GAN architectures created for the generation of diverse and non-deterministic outputs are being considered.

REFERENCES

[1] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, “Deep learning and process understanding for data-driven earth system science,” *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[2] T. Schneider, S. Lan, A. Stuart, and J. Teixeira, “Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations,” *Geophysical Research Letters*, 2017.

[3] R. Essery, N. Rutter, J. Pomeroy, R. Baxter, M. St”ahli, D. Gustafsson, A. Barr, P. Bartlett, and K. Elder, “Snowmip2: An evaluation of forest snow process simulations,” *Bull. Amer. Meteor. Soc*, vol. 90, pp. 1120–1136, 2009.

[4] A. M. Rhoades, A. D. Jones, and P. A. Ullrich, *Assessing Mountains as Natural Reservoirs with a Multi-Metric Framework*. Earth’s Future (In Revision), 2018.

[5] M. S. Raleigh, J. D. Lundquist, and M. P. Clark, “Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, hydrol,” *Earth Syst. Sci*, vol. 19, pp. 3153–3179, 2015.

[6] K. S. Jennings, T. S. Winchell, B. Livneh, and N. P. Molotch, *Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere*. Nat. Commun, 2018.

[7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” *ArXiv e-prints*, Nov. 2016.

[8] B. Livneh, T. J. Bohn, D. W. Pierce, F. Munoz-Arriola, B. Nijssen, R. Vose, D. R. Cayan, and L. Brekke, “A spatially comprehensive, hydrometeorological data set for mexico, the u.s., and southern canada 19502013,” *Scientific Data*, vol. 2, no. 150042, 2015.

[9] D. Hall, J. Stewart, and C. Tierney, “Adversarial networks for satellite simulation and dataset translations,” *Proceedings of the 8^{th} Workshop on Climate Informatics, Boulder, CO, 2018*, 2018.