MODEL AND DATA UNCERTAINTY FOR SATELLITE TIME SERIES FORECASTING WITH DEEP RECURRENT MODELS

Marc Rußwurm[1], Mohsin Ali[2], Xiao Xiang Zhu[2,3], Yarin Gal[4], Marco Körner[1]

[1]Chair of Remote Sensing Technology
Department of Aerosapce and Geod.
Technical University of Munich

[2]Department of EO Data Science
Remote Sens. Technology Institute
German Aerospace Center

[3]Signal Processing in Earth Obs.
Department of Aerosapce and Geod.
Technical University of Munich

[4]OATML Research Group
Department of Computer Science
University of Oxford

ABSTRACT

Deep Learning is often criticized as being a black-box method that provides accurate predictions, but a limited explanation of the underlying processes and no indication when to not trust those predictions. Equipping existing deep learning models with an (general) notion of uncertainty can help mitigate both these issues. The Bayesian deep learning community has developed model-agnostic methodology to estimate both data and model uncertainty that can be implemented on top of existing deep learning models. In this work, we test this methodology for deep recurrent satellite time series forecasting and test its assumptions on data and model uncertainty. We tested its effectiveness on an application on climate change where the activity of seasonal vegetation decreased over multiple years.

Index Terms— Deep Learning, Uncertainty, Recurrent Neural Networks, Satellite Time Series, Forecasting, Climate

1. INTRODUCTION

The Earth's surface is governed by spatio-temporal processes that are monitored by satellite sensors at regular temporal intervals. Developing models for these complex processes and making use of the large quantity of satellite data is a central objective of today's remote sensing research. Deep learning models approximate complex functions without requiring explicit labels by predicting the next time instance in a sequence of observations. This can be used to obtain an initial model parameterization from unlabeled data before fine-tuning on specific tasks, as commonly done in natural language processing, or to forecast future time instances. Modeling and forecasting future states with a model is a key tool for detecting deviations from expected behavior, such as detecting anomalies, or identifying drifts in climate and vegetation. No measurement, however, is without error and forecastings future time instances accumulates uncertainty of the prediction. Allowing us to assess the certainty of prediction of a deep learning model can be useful in providing insights for the understanding of complex long-term model forecasts.
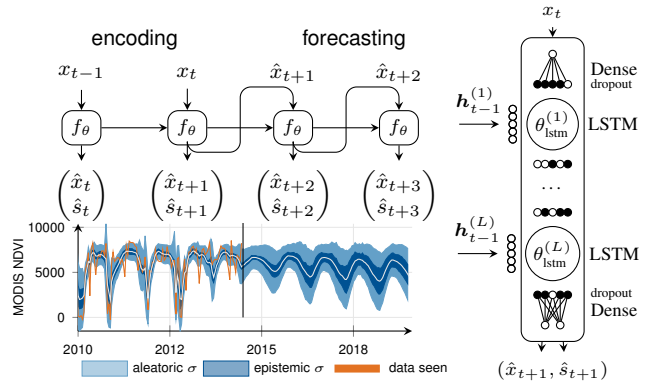

(a) Auto-regressive time series forecasting under uncertainty. Schematic and example.

(b) Recurrent model f_θ used in this work.

Fig. 1: Schematic illustration of the autoregressive time series forecasting process and its model f_θ.

2. RELATED WORK

Work on modeling processes on the Earth surface from satellite time series has been predominantly focused on analysis rather than forecasting. Eklundh and Jönsson [1], for instance, fit asymmetric Gaussian functions to time series and derive meaningful parameters from these curves. Other analysis methods, like harmonic regression, as in the BFAST (Breaks For Additive Season and Trend) [2] algorithm, or Fourier series, sometimes known as HANT (Harmonic ANalysis of Time Series)[3], can be used to extrapolate models into future time instances. Similarly, the CCDC [4] algorithm, fits a harmonic function with a robust iteratively reweighted least squares optimization to satellite time series and utilized all available Landsat data. Models based on Auto Regressive Integrated Moving Average (ARIMA) have been employed for forecasting drought events [5]. Deep recurrent neural networks have been recently come into focus of the Earth system science community [6] to model temporal processes. All of these models aim at functionally approximating the processes on the Earth's surface sensed remotely by satellites with uncertainty in measurements, *i.e.*, the *aleatoric uncertainty*. Also, these models often don't provide a notion of
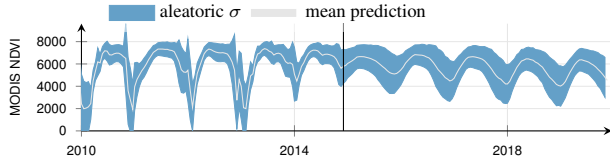
(a) aleatoric data uncertainty
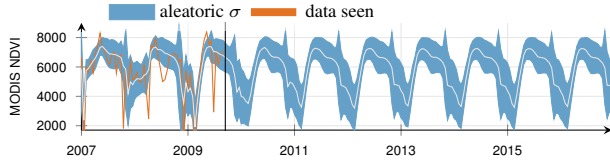
(b) aleatoric uncertainty with unsmoothed data
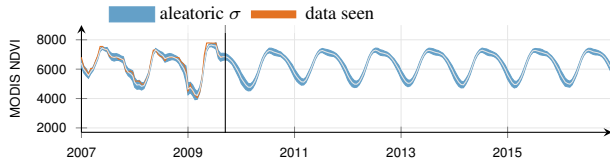
(c) aleatoric uncertainty with data smoothed with median-3 filter

Fig. 2: Examples of aleatoric (Fig. 2a) data-uncertainty. Aleatoric (data) uncertainty reduces when less noise is within the data. Compare the original data in Fig. 2b with prediction of median-smoothed data in Fig. 2c.

(a) epistemic model uncertainty

(b) epistemic uncertainty with injected data

(c) combined model uncertainty

Fig. 3: Examples of epistemic (Fig. 3a) model-uncertainty combined to a joint uncertainty model (Fig. 3c). Epistemic (model) uncertainty reduces when observing data, but increases when forecasting new time points.

confidence, *i.e.*, the *epistemic uncertainty*, which becomes increasingly relevant when using the learned model to forecast future states in a long-term context where model errors are expected to accumulate pushing the model to predict on states it never saw at training time. Recently, two approaches to model aleatoric [7] and epistemic uncertainties [8] have been developed in the Bayesian deep learning community. Kendall & Gal [9] unified these in a common framework with examples from semantic segmentation in computer vision. The estimation of aleatoric uncertainty can also be employed for multitask learning [10].
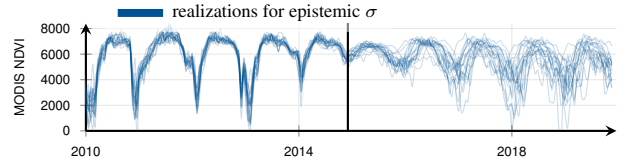
3. METHOD

Given a sequence of T data points $\boldsymbol{x} = (x_0, \ldots, x_T)$, we aim at forecasting N future time points $\hat{x}_{T+1}, \ldots, \hat{x}_{T+N}$. A recurrent neural network model
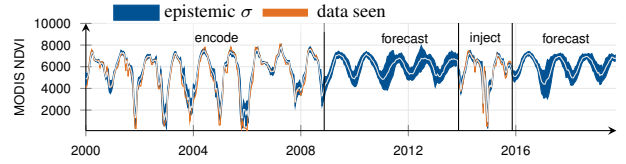
$$\hat{x}_{t+1} = f_\theta(x_t, x_{t-1}, \ldots, x_0) = f'_\theta(x_t, \boldsymbol{h}_{t-1}) \quad (1)$$

to make a prediction $\hat{x}_{t+1}$ of the next time point from an input x_t conditioned on a memory vector $\boldsymbol{h}_{t-1}$ that encodes features from previous data points $(x_{t-1}, x_{t-2}, \ldots)$. In the training phase, the optimal model parameters
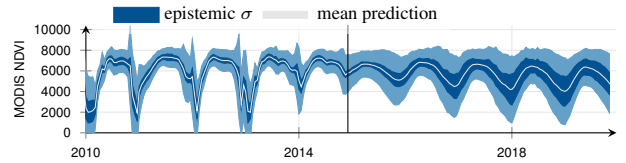
$$\theta^* = \arg\min_\theta \sum_{t=0}^{T} \mathcal{L}(x_t, f_\theta(x_{t-1}, x_{t-2}, \ldots)) \quad (2)$$

are optimized to predict the next time instance x_t. This is realized by minimizing an objective function $\mathcal{L}(\boldsymbol{x}, \theta)$ via minibatch stochastic gradient descent with samples from the training set $\boldsymbol{x} \in \mathcal{D}_{\text{train}}$. The model optimizes the parameters on the characteristics of the dataset and, for instance, internalizes seasonal patterns from the dataset. We illustrate this encoding of schematically in Fig. 1a. After encoding the sequence until time t, no measurements are available and we introduce the estimated future data points $\hat{x}_{t+1}$ to the model instead, as shown in Fig. 1a. This iterative auto-regressive forecasting process can be repeated for an arbitrary number of steps leading estimates of future time frames.

3.1. Recurrent Neural Network Model

The recurrent forecasting model is schematically shown in Fig. 1b. The input x_t at each time t is first mapped to a H-dimensional feature vector by a dense layer with ReLU activation function. L stacked Long Short-Term Memory [11] layers then combine this feature vector with features extracted from previous time frames via $\boldsymbol{h}_{t-1}$. A final dense layer maps the feature space to predict the future time instance $\hat{x}_{t+1}$ as well as the log variance $\hat{s}_{t+1}$. Dropout with 50% probability is applied before and after the encoding with the LSTM layers, as indicated by black nodes in Fig. 1.

3.2. Modeling Epistemic and Aleatoric Uncertainty

Modeling uncertainty is useful for interpretation of predicted future time instances. There are two types of uncertainties that have been unified in Kendall & Gal (2017)[9].

3.2.1. Aleatoric Uncertainty

Aleatoric data uncertainty describes the variability in the observed data points [9]. This can be homoscedastic, *i.e.*, constant noise for all observations, or heteroscedastic, *i.e.*, variable noise at each observation. Here, we model the heteroscedastic uncertainty by a model that predicts a second output for the log variance $\hat{s}_t := \log(\sigma_t^2)$ [7] of the data additionally to the prediction $[\hat{x}_t, \hat{s}_t] = f_\theta(x_{t-1}, x_{t-2}, \dots)$ at each time instance t. The two model outputs are balanced by a joint objective function

$$\mathcal{L}(\boldsymbol{x}, \theta) = \frac{1}{T} \sum_{t=0}^{T} \frac{1}{2} \exp(-\hat{s}_t)(x_t - \hat{x}_t)^2 + \frac{1}{2}\hat{s}_t \quad (3)$$

that combines the mean squared error $\frac{1}{T} \sum_{t=0}^{T}(x_t - \hat{x}_t)^2$ between predicted next time instance $\hat{x}_t$ and ground truth x_t with the inverse of the aleatoric variance $\exp(-\hat{s}_t) = \exp(-\log(\hat{\sigma}_t^2)) = \hat{\sigma}_t^{-2}$. The regularization term $\frac{1}{2}\hat{s}_t$ prevents the model from predicting $\hat{s}_i \rightarrow \infty$ which would lead to trivial zero loss. Data points were normalized with mean and standard deviation over time series of the dataset. We show an example of aleatoric uncertainty in Fig. 2a.

3.2.2. Epistemic Uncertainty

Epistemic model uncertainty increases when extrapolating beyond seen data. It decreases when new data points have been observed [9], as visible in Figs. 3b and 3c. It is modeled by performing K model forward passes with dropping out single rows in the weight matrix $\hat{\theta}_i \sim p(\theta)$ with $p_{\text{drop}} = 0.5$ dropout probability [8]. In consequence, we draw a model $f_{\hat{\theta}_i} \sim p(f_\theta)$ from a model family f_θ [8]. Note that dropout is only applied in the test-phase. We show an example of 20 draws in Fig. 3a. With K predictions at each time t, we can approximate the combined variance [9] as

$$\mathbb{V}[\hat{x}_t] \approx \underbrace{\frac{1}{K} \sum_{i=0}^{K} \hat{x}_{i,t}^2 - \left(\frac{1}{K} \sum_{i=0}^{K} \hat{x}_{i,t}\right)^2}_{\text{epistemic}} + \underbrace{\frac{1}{K} \sum_{i=0}^{K} \hat{\sigma}_{i,t}^2}_{\text{aleatoric}} \quad (4)$$

which is the sum of epistemic variances and the mean aleatoric variance of all predictions, as shown in Fig. 3c.

4. EXPERIMENTS

We investigated the behavior of aleatoric (data) and epistemic (model) uncertainties on seasonal vegetation dynamics.
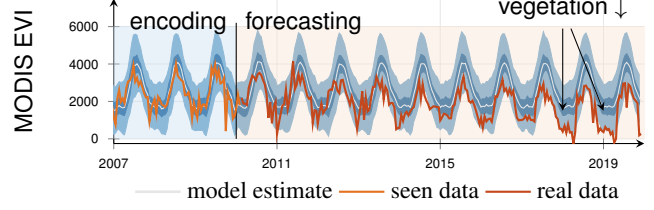

Fig. 4: Decreasing vegetational activity in Canada.

These are obtained from the normalized difference vegetation index (NDVI) measured from MODIS-Terra. A set of 1000 time series with 460 observations from 2000 to 2020 at random locations within an area of interest in Bavaria, Germany, have been obtained via Google Earth Engine. This model is then shown a time series from 2000 to 2015 from the test dataset and forecasts data until 2020.

Aleatoric Uncertainty. We can observe in Fig. 2b that the aleatoric uncertainty remained comparatively constant throughout the forecasting. Since it captures the data uncertainty, no increase in data or measurement error is expected at future time steps. When we artificially reduced the data noise by filtering the time series by a 5-timeframe median filter, as in Fig. 2c, we observed that the estimated aleatoric uncertainty reduced accordingly.

Epistemic Uncertainty. Epistemic model-uncertainty is realized by Monte Carlo dropout sampling [8], which can be interpreted as sampling models from a model family with different weights dropped out, as described in Section 3.2.2. An example of realizations can be seen in Fig. 3a. When forecasting future time instances, the different model realizations diverge with absence of new common observations. This intuition was confirmed in Fig. 3b where we first forecasted several data points from 2009 to 2014 and the epistemic uncertainty increased. Then, we *injected* new data points between 2014 and 2016. Now, all model realizations were provided with the same measured data points. Hence, all model realizations converged to a more similar prediction which lead to reduced epistemic model uncertainty following Eq. (4). Throughout this process, the aleatoric model uncertainty remained comparatively constant.

5. APPLICATIONS

In this section, we show one Earth observation application where the uncertainty of the predictions provides relevant feedback on the significant deviations of the expected behavior. Canada dataset have been acquired similarly to the dataset gathered over Germany from the previous section, but show results on the *Enhanced Vegetation Index*. For this dataset, we acquired 1000 time series from 2000 to 2010 and subdivided them into training, validation, and testing partitions at a ratio of 3:1:1. For this experiment, we trained the model on time series from this region up to 2010 and forecasted the vegetation dynamics in EVI until 2020. Hence, we modeled

the expected vegetation behavior with quantifiable uncertainty and compared it to the observed measurements from the MODIS satellite.We introduced the day of year of the observation t_{doy} as two additional features $\sin(t_{\text{doy}}), \cos(t_{\text{doy}})$ to the model. This feature introduces additional design knowledge on the seasonality of the data which helps the model to reduce its epistemic uncertainty for long-term prediction by building structure into the model's predictive uncertainty.

Detection of Vegetation Drift and Climate Change. In Fig. 4, we focus on a region on the east coast of Canada. The actual MODIS observations since 2010 are drawn in red and reveal a diverging trend from the vegetation model trained on data from 2000 to 2010. Similar to the previous experiment, we can make decisions with uncertainty bounds on the deviation of the measured vegetation from the learned model that captured the natural variability of the region. For instance, we see that the measured vegetation activity has been systematically outside the 1σ confidence interval in the last two years of 2018 and 2019. Additionally, the model uncertainty provides a confidence interval hinting on when we cannot trust our model any more, at which point we might decide to collect new data to retrain the model.

6. DISCUSSION

It is encouraging to see that uncertainty calculation methods from the computer vision literature can work out-of-the-box for remote sensing time series data. Both aleatoric and epistemic uncertainties have shown expected behaviors, as evaluated in Section 4. The epistemic uncertainty increased as it forecasted farther points in the future. The aleatoric uncertainty was not affected since the variability of satellite reflectances does not change with future time points. The dataset in Germany that we utilized was homogenious in nature and without irregularities. In future work, we would like to evaluate these accuracy metrics on more heterogeneous remote sensing time-series use-cases. Quantifying uncertainties on anomaly events, as, for instance, induced by natural disasters, would be a challenging, but interesting application of epistemic and aleatoric uncertainties. The recurrent neural network did learn a general notion of seasonality from the data. This, however, can be done similarly by decomposing the original time series into linear components that capture trend and seasonality. Incorporating the prior knowledge of seasonality in the model by letting the model predict only the remainder rather than the entire time series may lead to better forecasting performance. Also, we z-normalized the data by mean and standard deviation. This limited the models to stationary time series that have no trend in the data. Detrending the time series before forecasting or predicting differences rather than absolute future time points may be a suitable adaptation for non-stationary time series that we want to explore in the future.

It is also essential to compare the current methods with other uncertainty methods, such as Gaussian processes or ensemble methods, to identify the most promising ones for a remote sensing use-case.

7. CONCLUSION

In this work, we analyzed aleatoric and epistemic uncertainties for the application of satellite time series, a needed but understudied topic. After describing the calculation of uncertainties in the context of time series modeling, we presented experiments that demonstrated the characteristics and differences of aleatoric and epistemic uncertainties on MODIS satellite time series for vegetation modeling and forecasting. We presented results on an example application which showed the potential of estimating uncertainties for Earth observation and remote sensing tasks by learning a regular dynamic with uncertainty bounds to which observed data can be compared. Finally, we discussed the possible challenges and future work necessary to quantify uncertainties and emphasized the importance of evaluating the qualities of calculated uncertainties.[1]

8. REFERENCES

[1] Per Jönsson and Lars Eklundh, "Timesat—a program for analyzing time-series of satellite sensor data," *Computers & Geosciences*, vol. 30, no. 8, pp. 833–845, 2004.

[2] Jan Verbesselt, Rob Hyndman, Achim Zeileis, and Darius Culvenor, "Phenological change detection while accounting for abrupt and gradual trends in satellite image time series," *Remote Sensing of Environment*, vol. 114, no. 12, pp. 2970–2980, 2010.

[3] Jie Zhou, Li Jia, and Massimo Menenti, "Reconstruction of global modis ndvi time series: Performance of harmonic analysis of time series (hants)," *Remote Sensing of Environment*, vol. 163, pp. 217–228, 2015.

[4] Zhe Zhu and Curtis E Woodcock, "Continuous change detection and classification of land cover using all available landsat data," *Remote sensing of Environment*, vol. 144, pp. 152–171, 2014.

[5] Ping Han, Peng Xin Wang, Shu Yu Zhang, and De Hai Zhu, "Drought forecasting based on the remote sensing data using arima models," *Mathematical and computer modelling*, vol. 51, no. 11-12, pp. 1398–1403, 2010.

[6] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, et al., "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195, 2019.

[7] Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla, "Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2017.

[8] Yarin Gal and Zoubin Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, 2016, pp. 1050–1059.

[9] Alex Kendall and Yarin Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," in *Advances in neural information processing systems*, 2017, pp. 5574–5584.

[10] Lukas Liebel and Marco Körner, "Auxiliary tasks in multi-task learning," *arXiv preprint arXiv:1805.06334*, 2018.

[11] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

---

[1]Code and data is available at http://github.com/marccoru/igarss2020_code