Watermarking Time Series Diffusion Models

Lucas Fatas

Research Question

How to implement a watermarking method for 2D time series diffusion models with an emphasis on balancing detectability and invisibility?

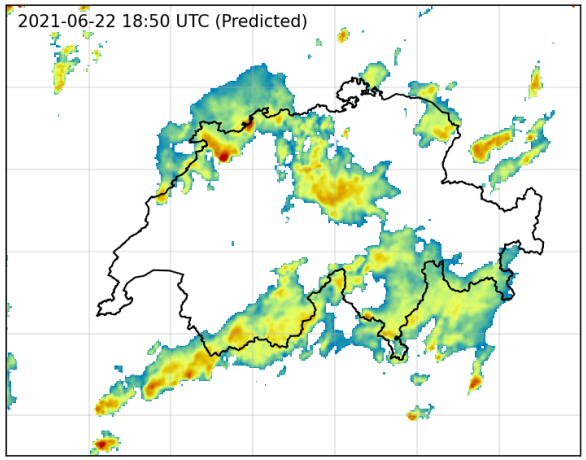
Background

- **Digital Watermarking:** Process of embedding a piece of code or a key in data in order for authenticity or ownership to be confirmed.
- **Current Watermarking Limitations:** Traditional watermarking methods focus on multimedia, leaving a gap in time series diffusion models.
- **Novel Approach:** Proposes a modification of the tree ring watermarking method tailored for the 2D time series model LDCast.
-

Methodology

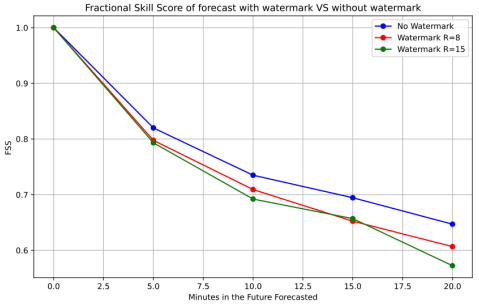
- **Model:** LDCast, a 2D time series model for precipitation forecasting
- **Watermarking Technique**: Adaptation of the tree ring watermarking technique from image-based models to time series models
- **Watermark Generation:** Watermarks were embedded into initial noise vector use by the model to generate the precipitation forecast
- **Watermark Detection:** Detection involved a reverse engineering process of the data to reconstruct the initial noise vector to confirm watermark presence.

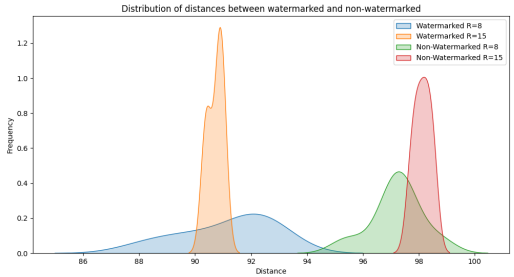
Example: Timestamp from LDCast data



Results

- **Experimental Setup:** Evaluating invisibility and detectability of a large watermark (R=15) and small watermark (R=5)
- **Evaluating Invisibility:** Measured by testing for the impact of watermark on the model's forecasting accuracy using the Fractional Skill Score (FSS)
- **Evaluating Detectability**: The effectiveness of watermark detection was evaluated by measuring the similarity between watermark embedded in the initial noise vector and the reconstructed noise vector of watermarked and non-watermarked data.





Discussion/Conclusion

- **Effectiveness of Watermark Size**: Larger watermark sizes (radius of 15) provided a better balance between detectability while still having minimal impact on the model's functionality
- **Successful Implementation:** The study successfully implemented a watermarking method for 2D time series diffusion models, proving the concept's feasibility.
- **Future Directions:** Suggested further testing for robustness against various attacks and exploring the application of this watermarking technique to other time series models.