Voyage from Observations to Solar Activity Forecast through AI/ML and Data Assimilation

Irina Kitiashvili

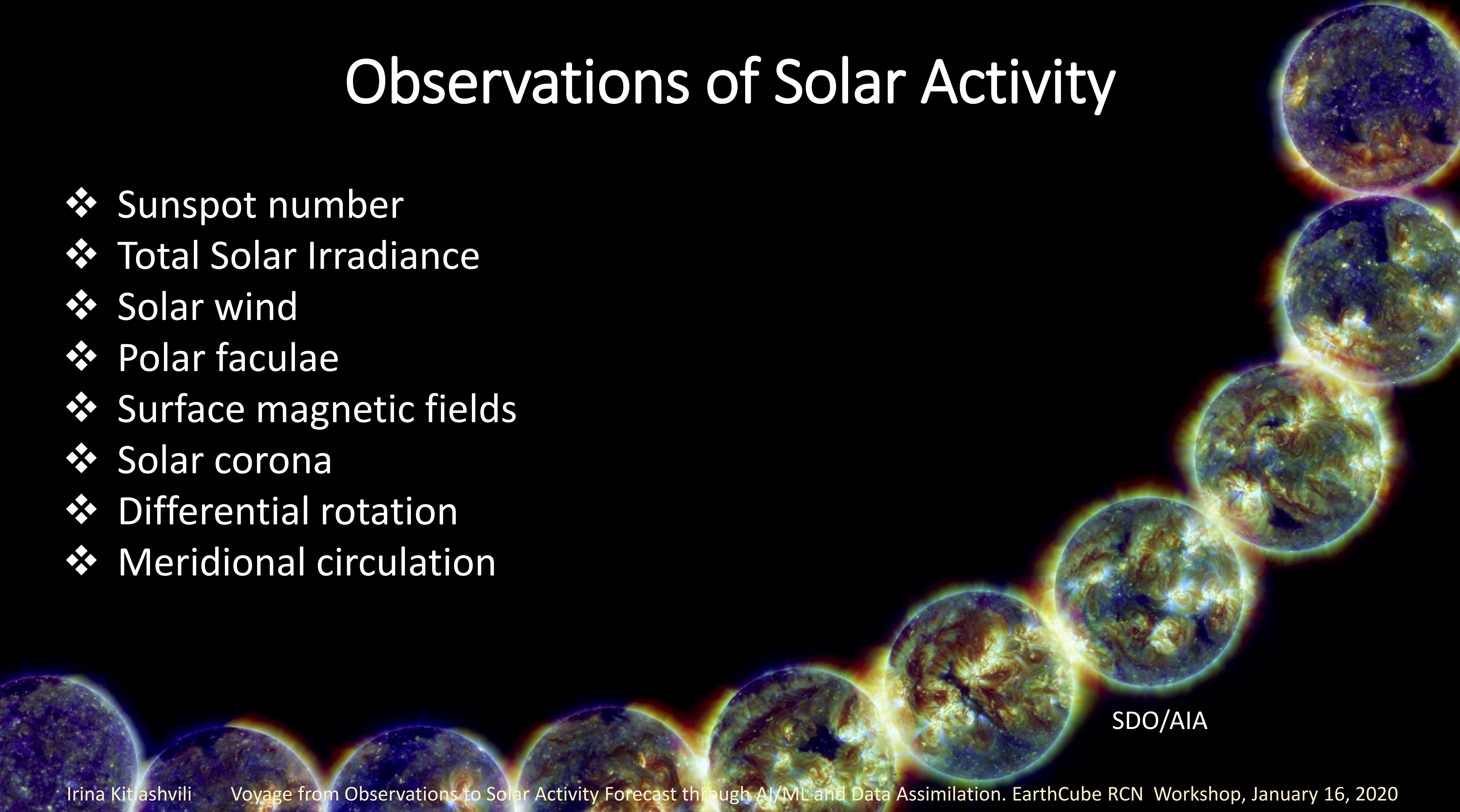
NASA Ames Research Center

Bay Area Research Environmental Institute

The research is funded by the NSF SHINE program AGS-1622341

Observations of Solar Activity

- Sunspot number
- Total Solar Irradiance
- Solar wind
- Polar faculae
- Surface magnetic fields
- Solar corona
- Differential rotation
- Meridional circulation

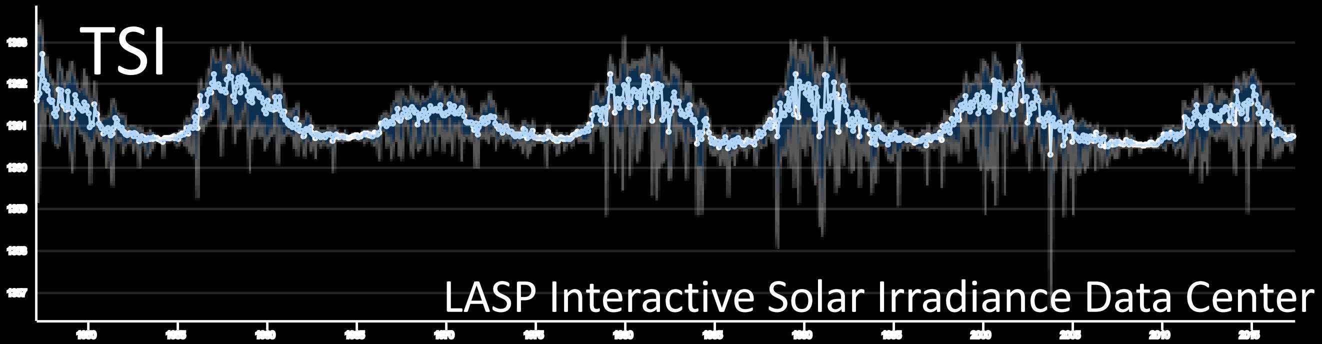


SDO/AIA

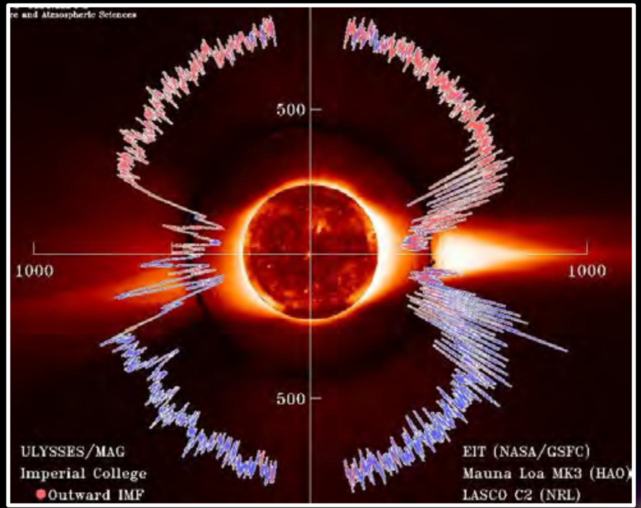
Observations of Solar Activity

1D observations

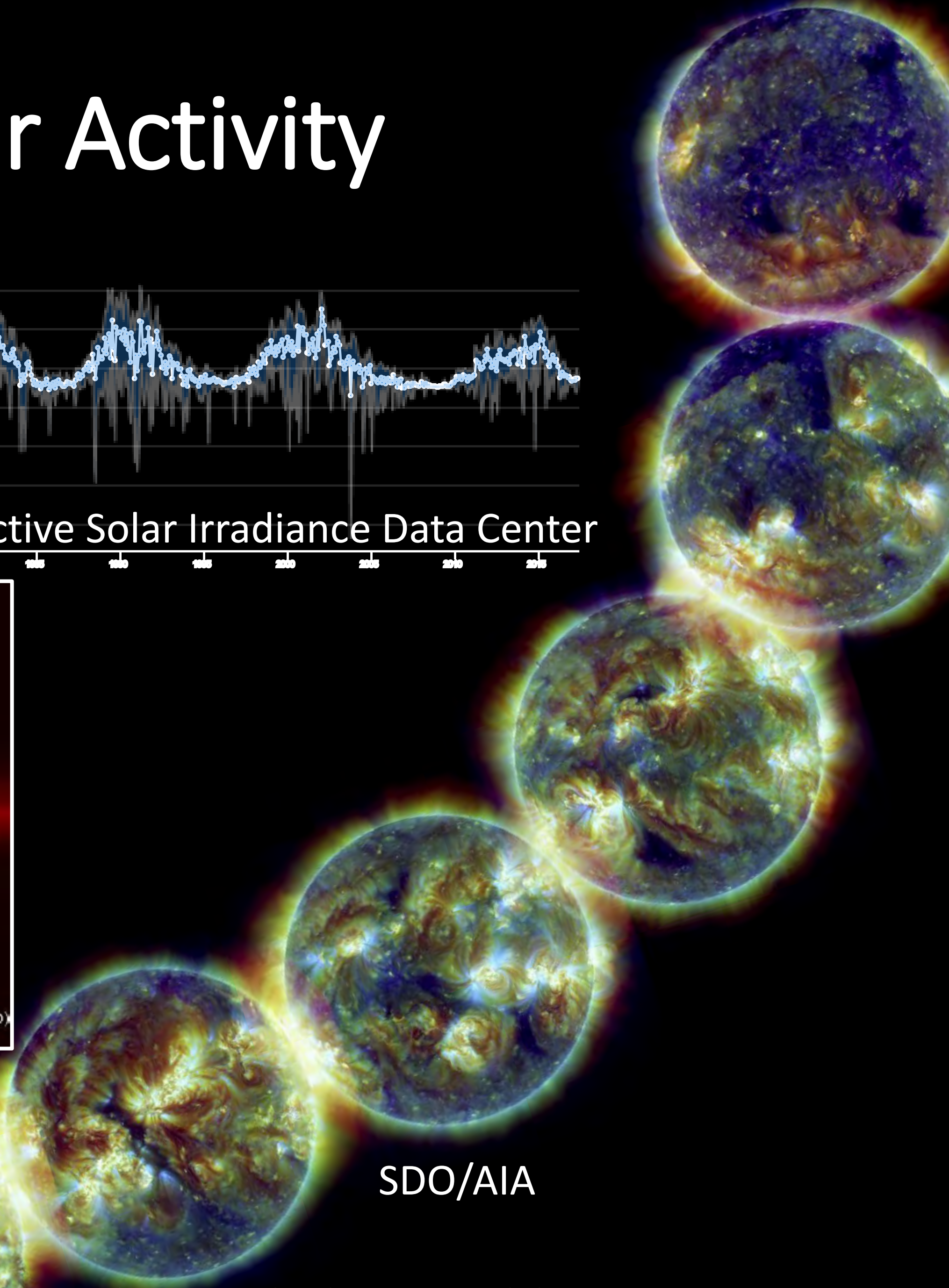
- Sunspot number
- Total Solar Irradiance
- Solar wind
- Polar faculae
- Surface magnetic fields
- Solar corona
- Differential rotation
- Meridional circulation



LASP Interactive Solar Irradiance Data Center



ULYSSES/SWOOPS



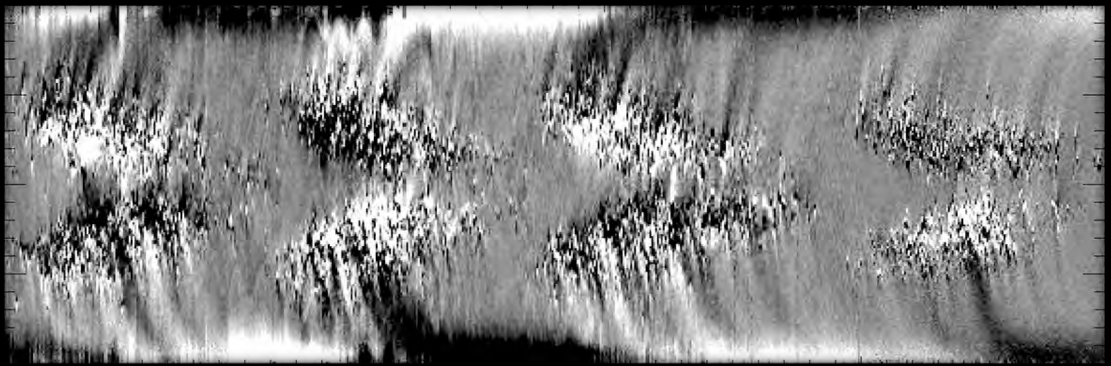
SDO/AIA

Observations of Solar Activity

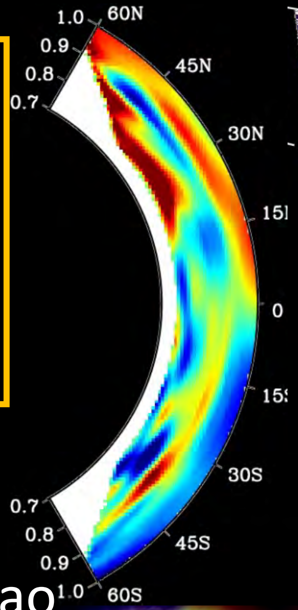
2D observations

- Sunspot number
- Total Solar Irradiance
- Solar wind
- Polar faculae
- Surface magnetic fields
- Solar corona
- Differential rotation
- Meridional circulation

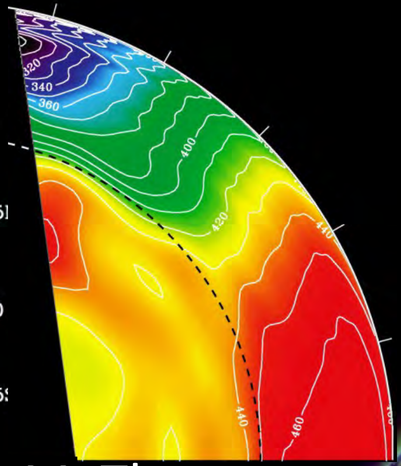
No subsurface magnetic field observations



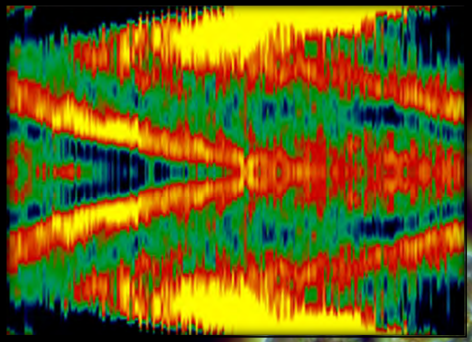
NSO
MDI/SoHO
HMI/SDO



J. Zhao



M. Thompson



R. Howe

SDO/AIA

Machine Learning vs Data Assimilation

Machine learning is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead.

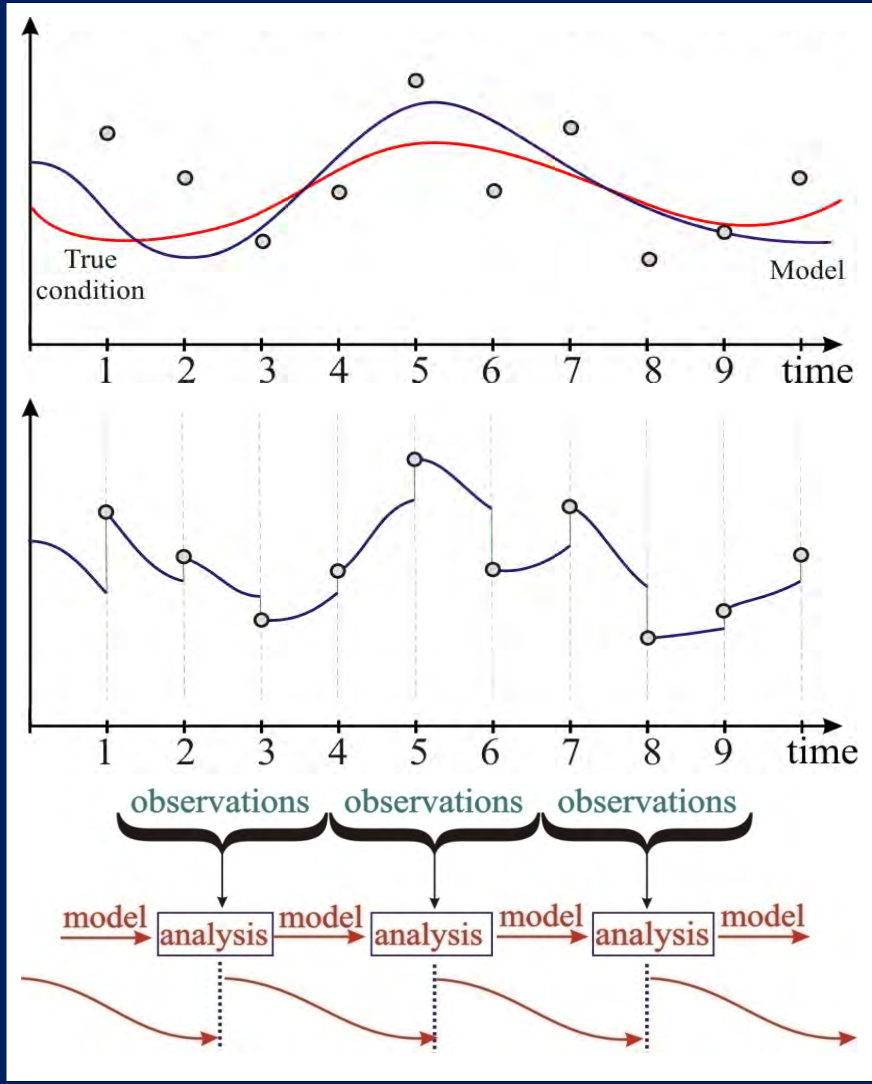
- Supervised learning
- Unsupervised learning
- Reinforcement learning
- Deep learning

Data assimilation is a mathematical discipline that seeks to optimally combine theory with observations.

- Wikipedia

- Sequential assimilation
  - Ensemble methods
  - Variational methods
- Non-sequential (retrospective) assimilation

Data Assimilation 101: Basic Concept & Kalman Filter



Observational data include errors, and a model constructed on their basis is characterized by some approximations; therefore, a prediction of the next set of observations will diverge from the real data.

$$d = M\psi^t + \varepsilon$$ observations

$$\frac{d\psi^t}{dt} = f\left(\psi^t, t\right) + q$$ model

$$\psi^f = M\psi^t + p^f$$ forecast

$$\psi^a = \psi^f + K\left(d - M\psi^f\right)$$ Best estimate of a state

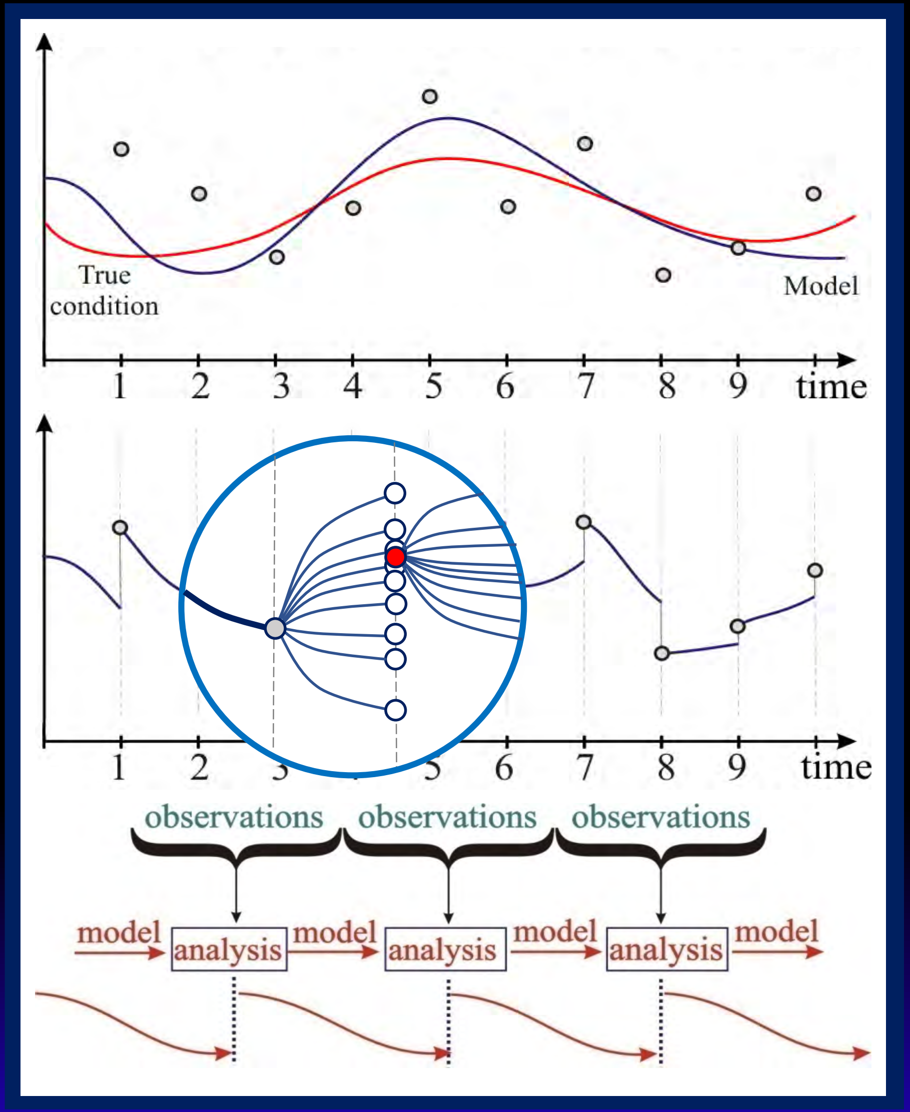
$$K = C_{\psi\psi}{}^f M^T \left(MC_{\psi\psi}{}^f M^T + C_{\varepsilon\varepsilon}\right)^{-1}$$ Kalman gain

$$J\left[\psi^a\right] = \left(\psi^f - \psi^a\right)\left(C_{\psi\psi}^f\right)^{-1}\left(\psi^f - \psi^a\right) +$$ Cost function

$$+\left(d - M\psi^a\right)^T W_{\varepsilon\varepsilon}\left(d - M\psi^a\right)$$

$W_{\varepsilon\varepsilon}$ is the inverse of the measurement error covariance matrix $C_{\varepsilon\varepsilon}$

Data Assimilation 101: Ensemble Kalman Filter



Observational data include errors, and a model constructed on their basis is characterized by some approximations; therefore, a prediction of the next set of observations will diverge from the real data.

$$d_j = M\psi^t + \varepsilon_j$$ observations

$$\frac{d\psi^t}{dt} = f\left(\psi^t, t\right) + q$$ model

$$\psi_j^f = M\psi^t + p_j^f$$ forecast

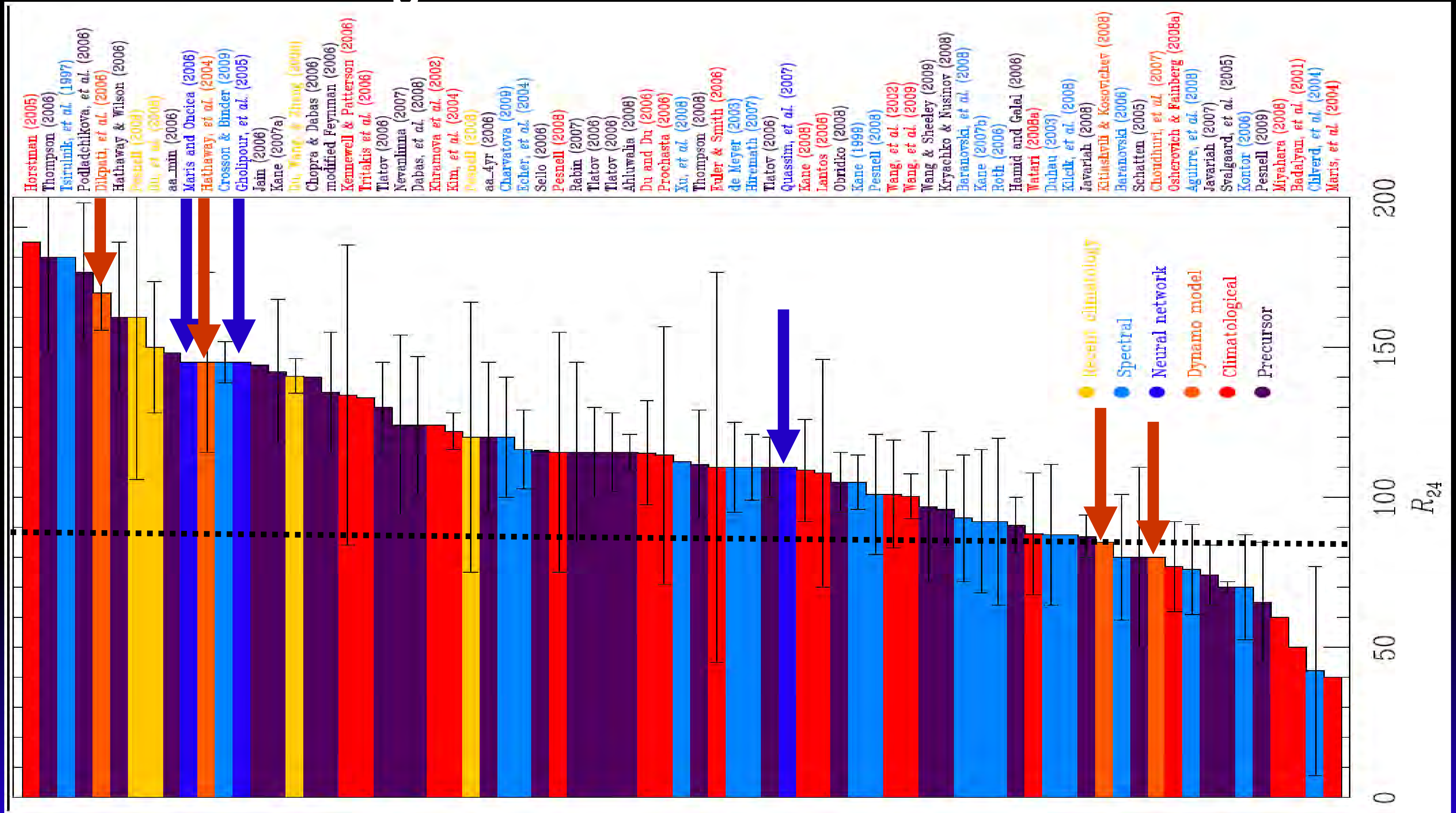
$$\psi_j{}^a = \psi_j{}^f + K\left(d_j - M\psi_j{}^f\right)$$ Best estimate of a state

$$K = \left(C_{\psi\psi}{}^e\right)^f M^T \left(M\left(C_{\psi\psi}{}^e\right)^f M^T + C_{\varepsilon\varepsilon}{}^e\right)^{-1}$$ Kalman gain

Methods of data assimilation allow us, with the help of the already constructed model and observational data, to determine the initial state that is in agreement with a new set of observations and to obtain a forecast of future observations and to estimate their errors.

Forecasts of Global Solar Activity: Solar Cycle 24

Machine Learning **Data Assimilation**

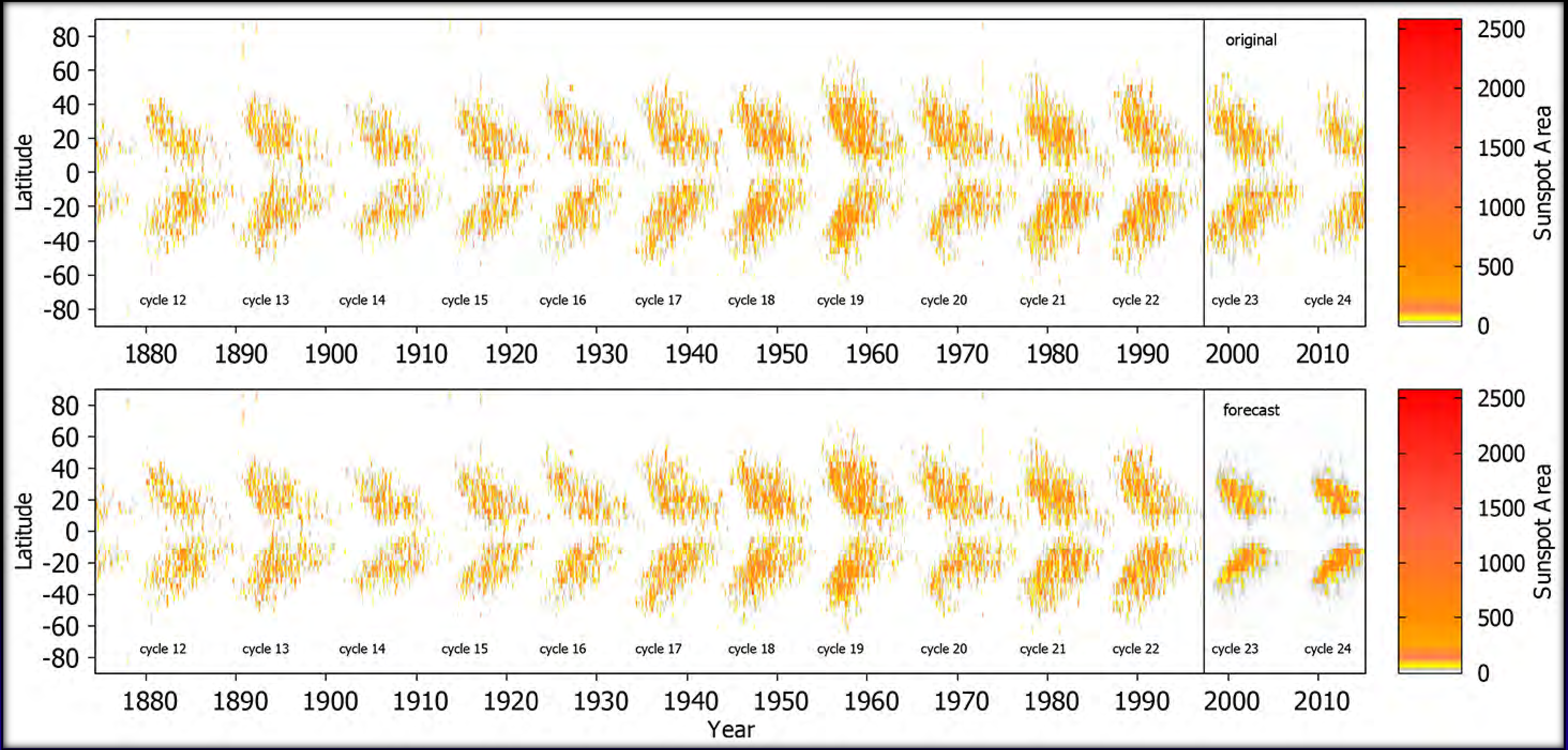


Pesnell, 2012

Forecast of Sunspot Area with a Deep Neural Network

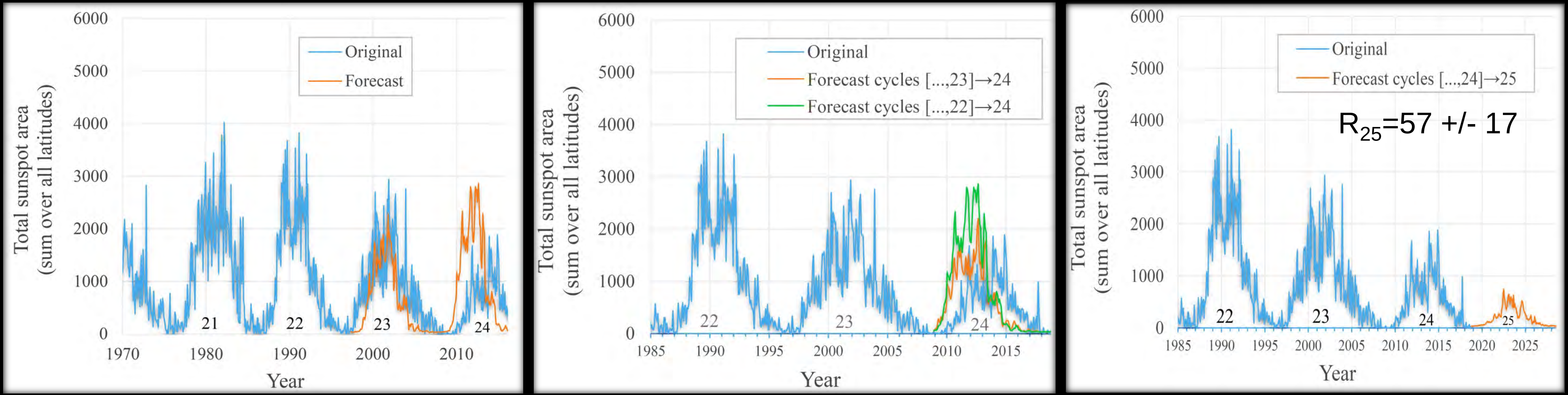
Base training set: 1646 latitudinal measurement times

Test set: 242 latitudinal measurement times



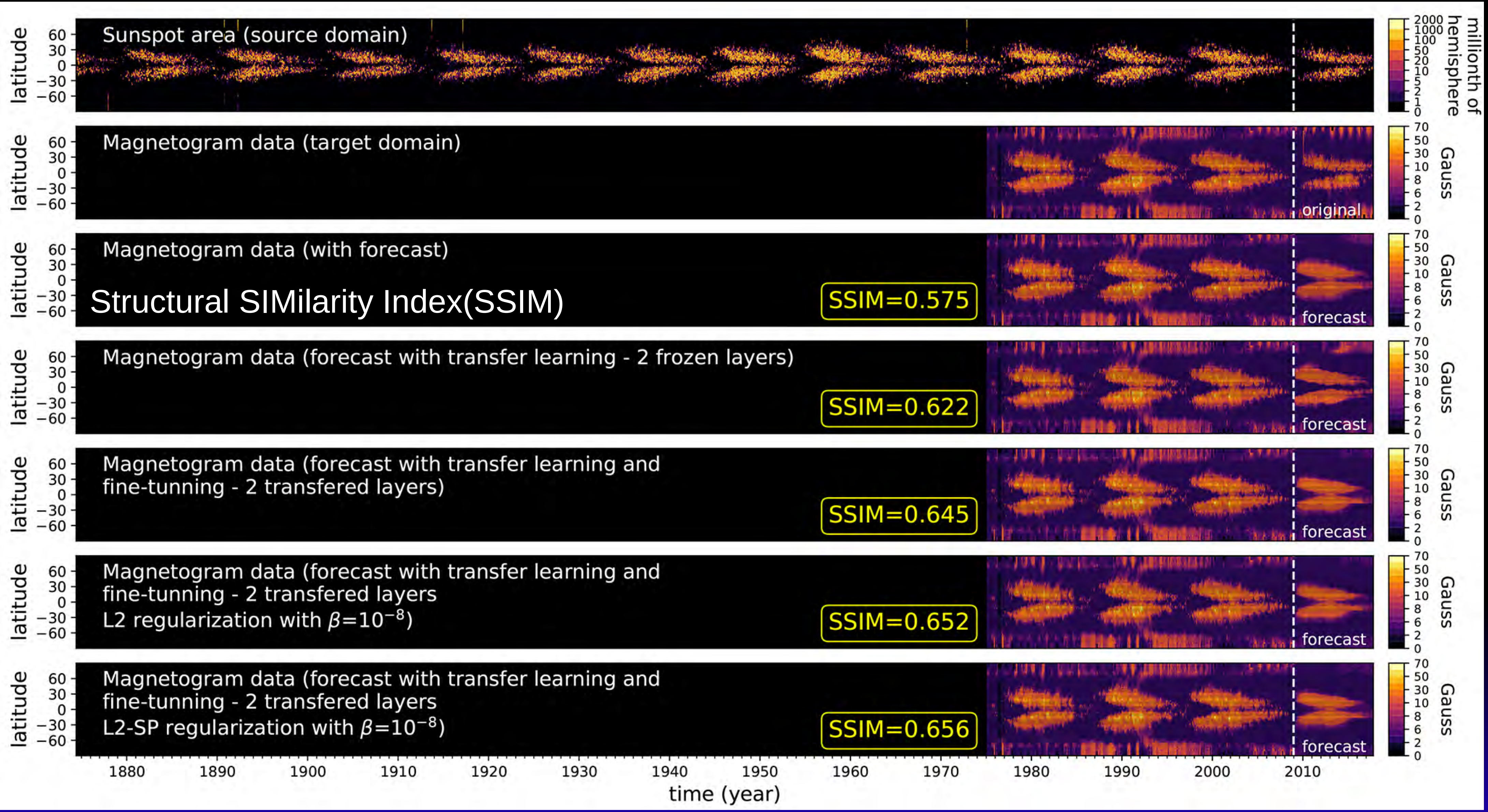
Covas et al., 2019

Solar Cycle predictions with a neural network



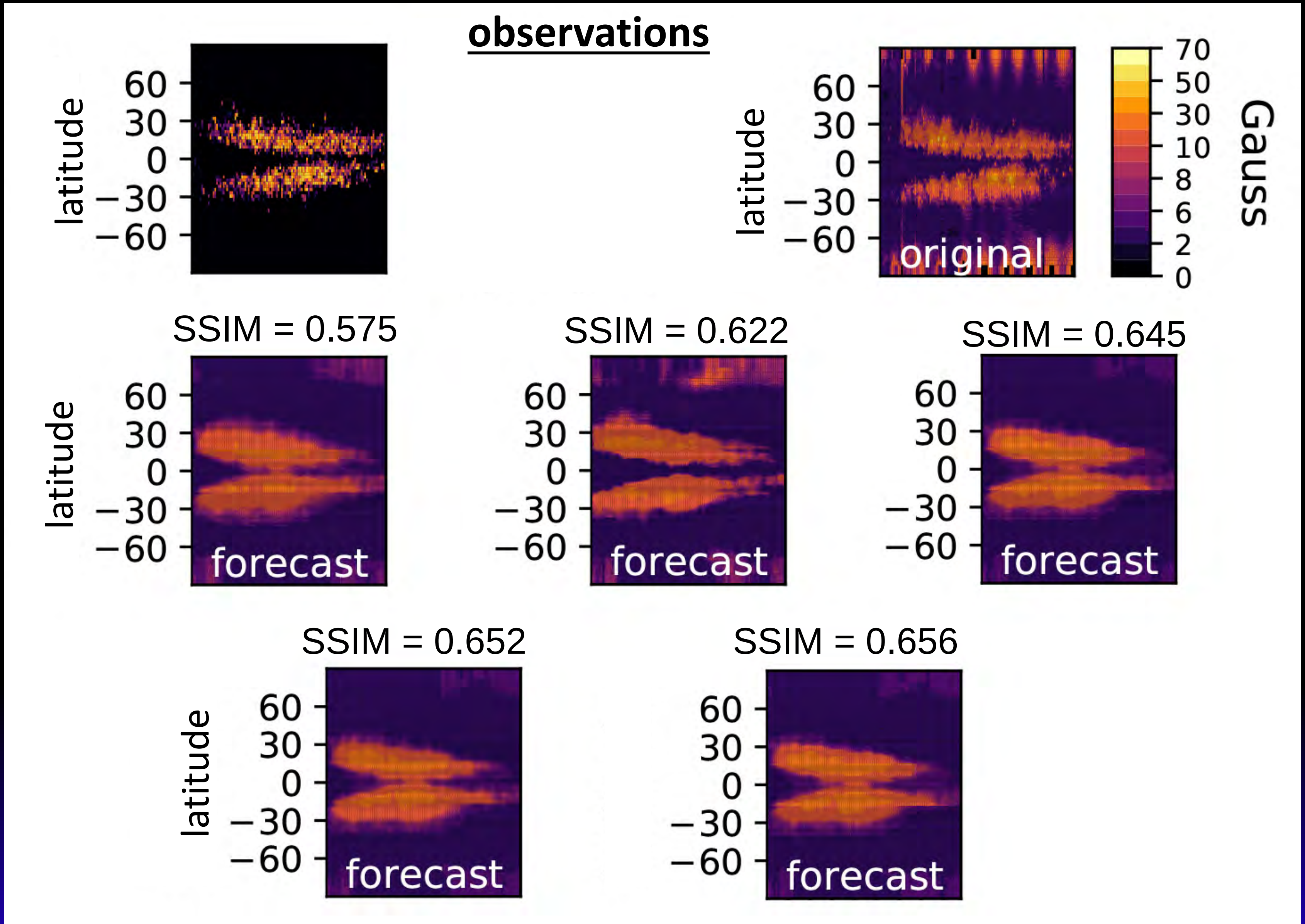
Covas et al., 2019

Magnetic field reconstruction with a neural network



Covas et al., 2019

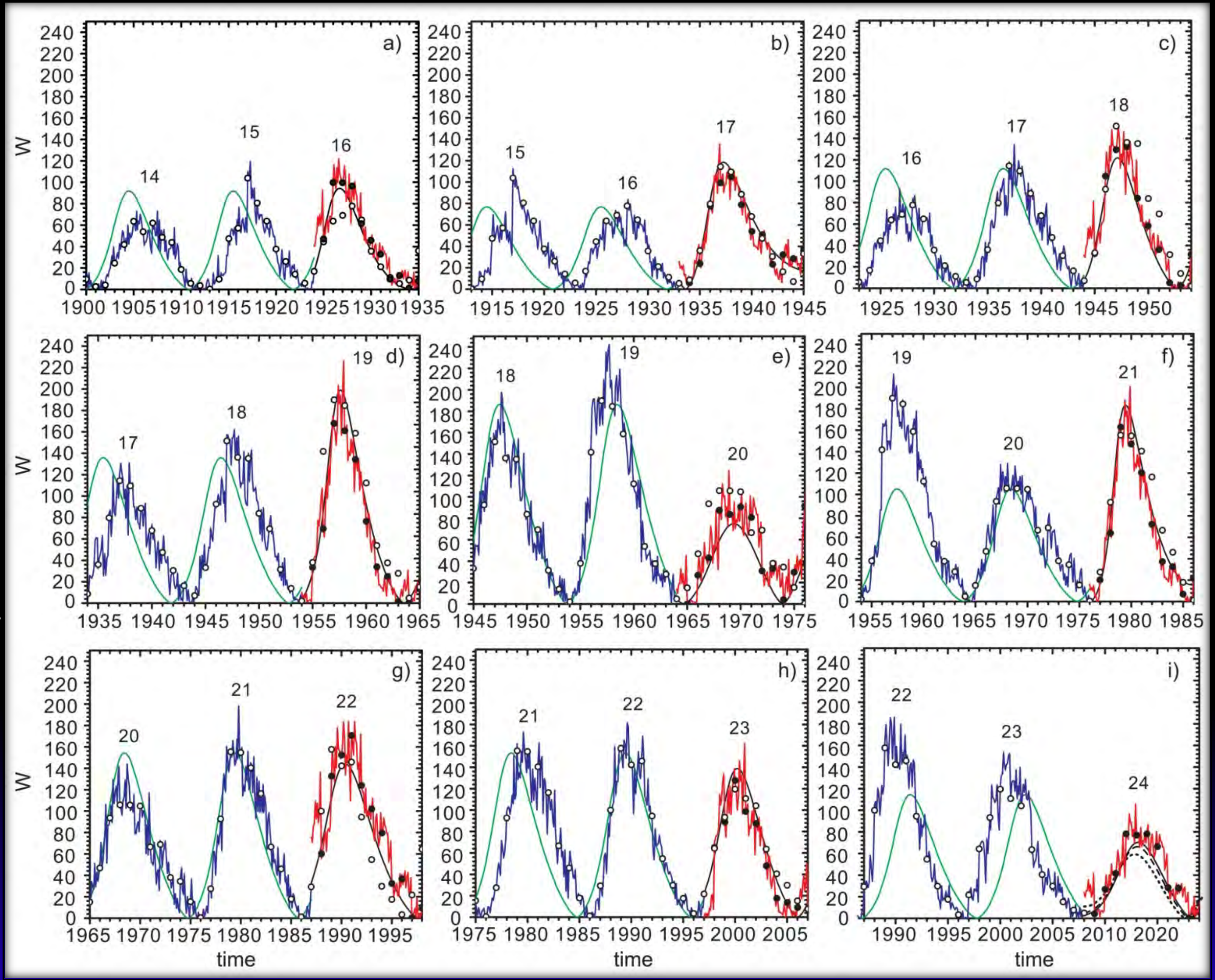
Magnetic field reconstruction with a neural network



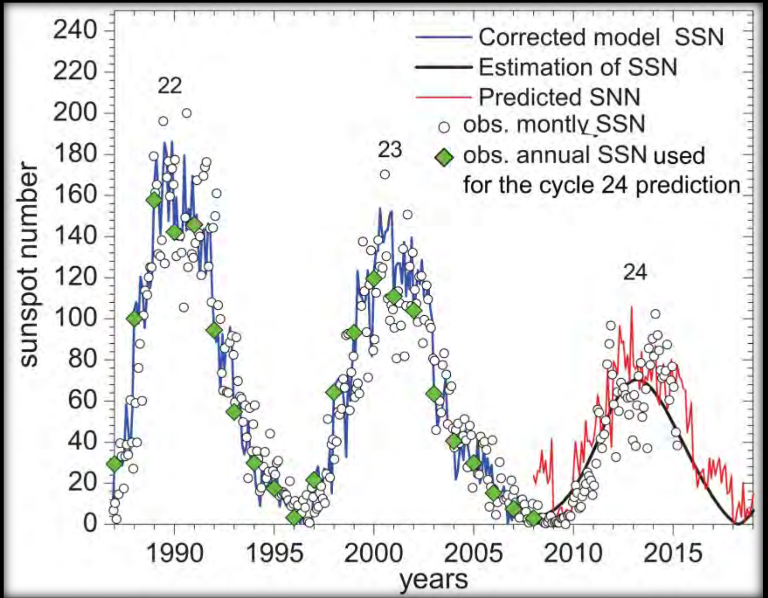
Covas et al., 2019

Forecast of solar activity with long sunspot number time-series

Data Assimilation

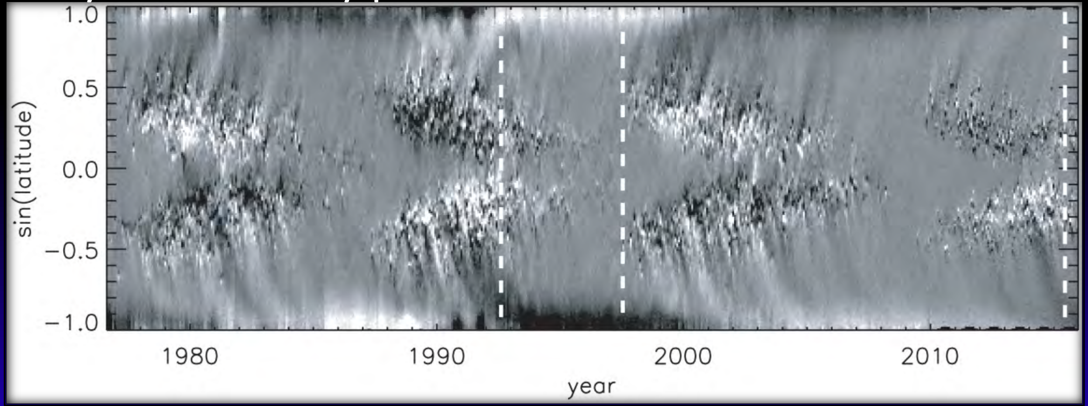


Kitiashvili & Kosovichev, 2008



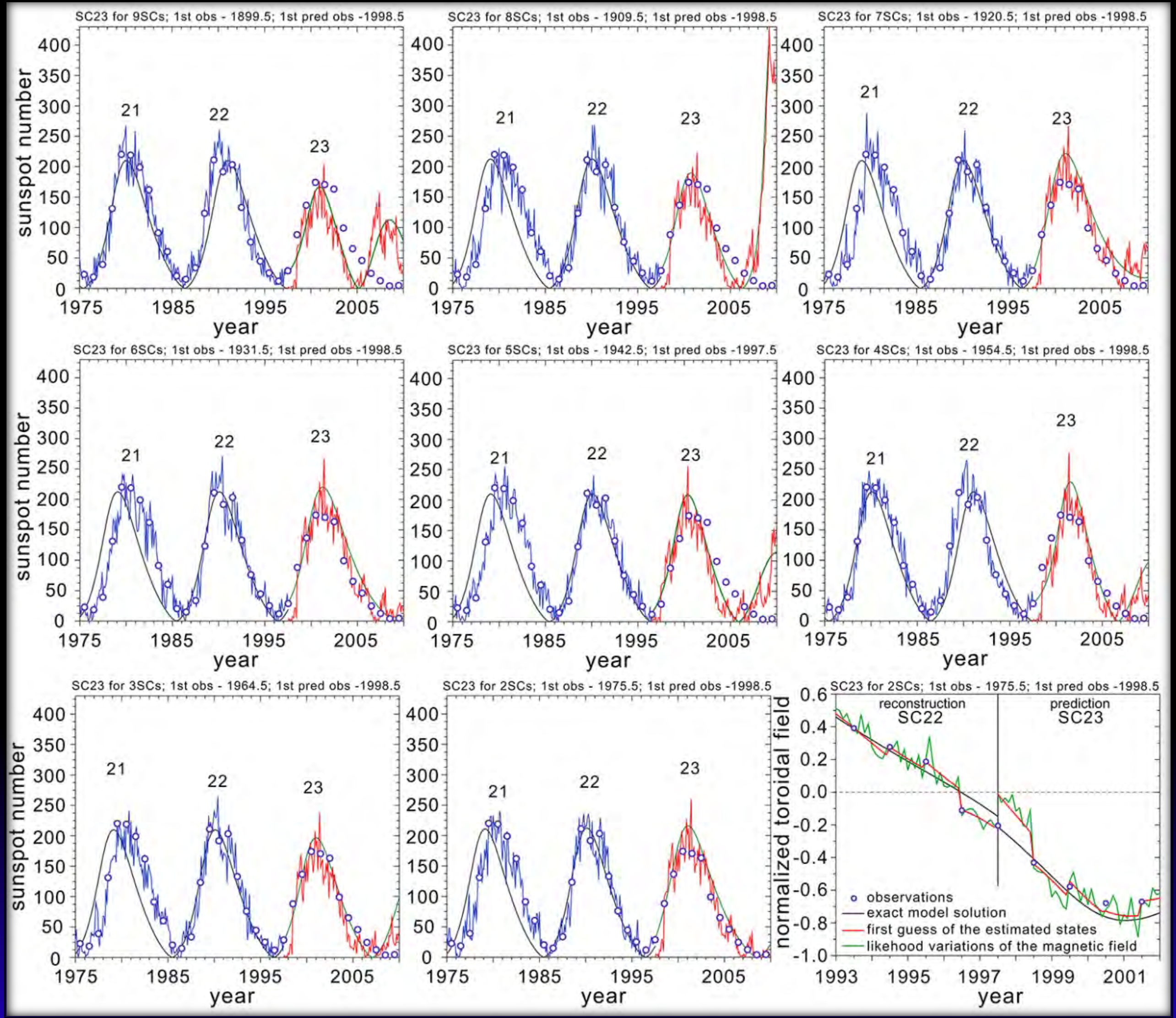
Kitiashvili, 2016

Early solar activity prediction



Forecast of solar activity with short sunspot number time-series

Data Assimilation



Kitiashvili, 2020

Criteria to identify an accurate model prediction:

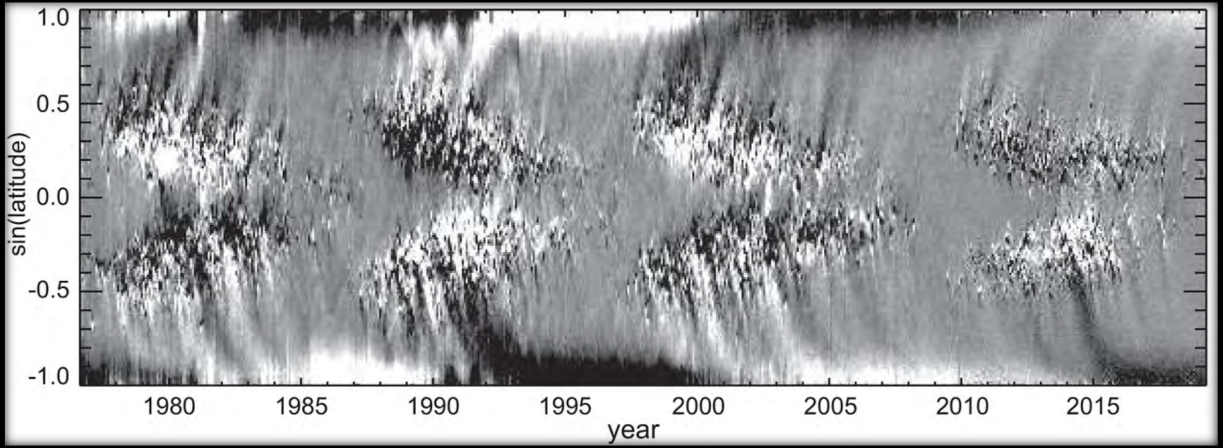
1) the signs of the last available observation (for toroidal field) and the corresponding model solution should be the same;
2) the exact model solution for the prediction phase must be consistent with the model solution for the reconstruction phase (no solution flattening, jumps, or 'bumps', but the solution may shift according to the new initial condition);
3) the corrected solution (first guess estimate) at the initial moment of time during the prediction phase should not be greater than the best-estimate variations of the toroidal field;
4) the phase discrepancy between the exact model solution and observations should not be greater than 2 years

Solar activity forecast using synoptic magnetograms

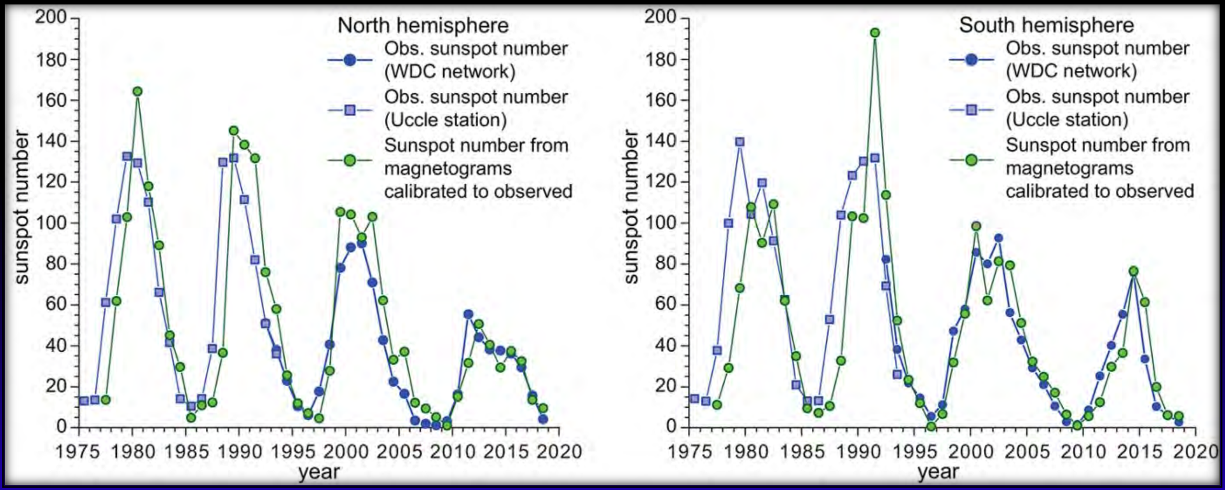
Data Assimilation

Synoptic magnetograms: SOLIS, MDI/SoHO, HMI/SDO

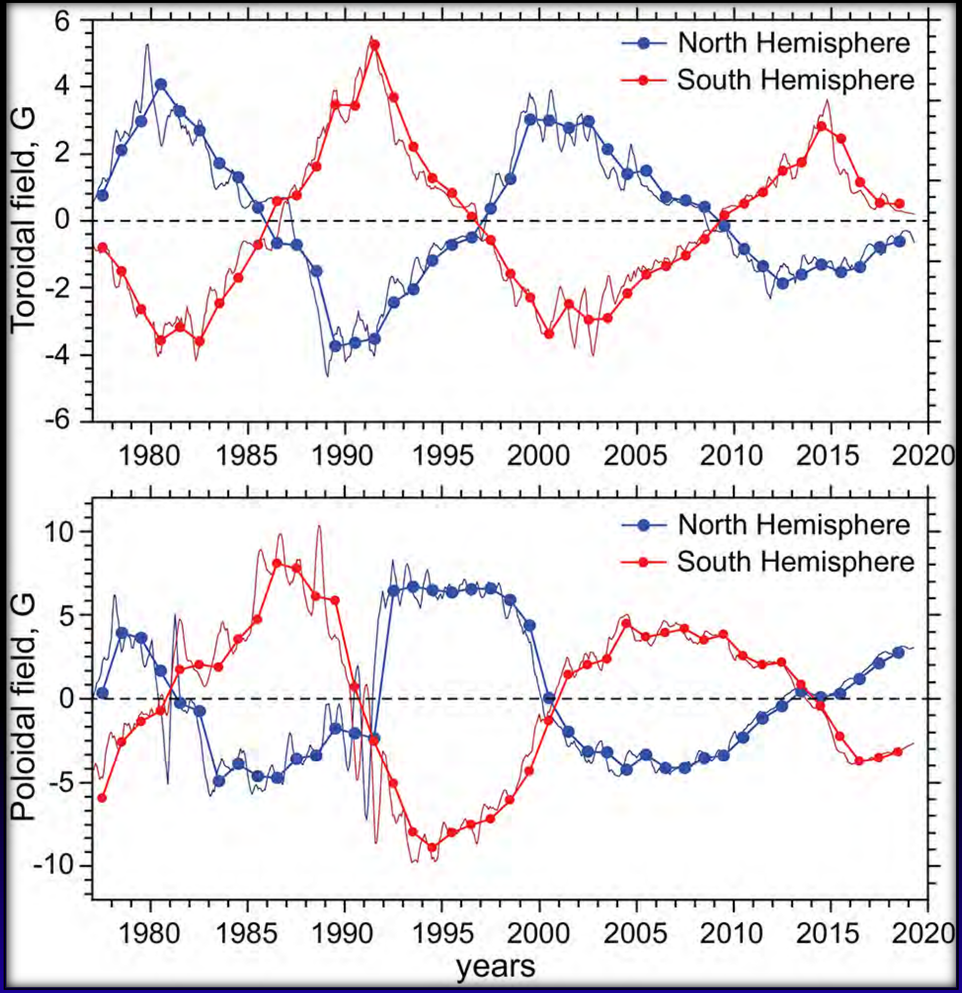
SC21 SC22 SC23 SC24



Sunspot number: observed vs estimated



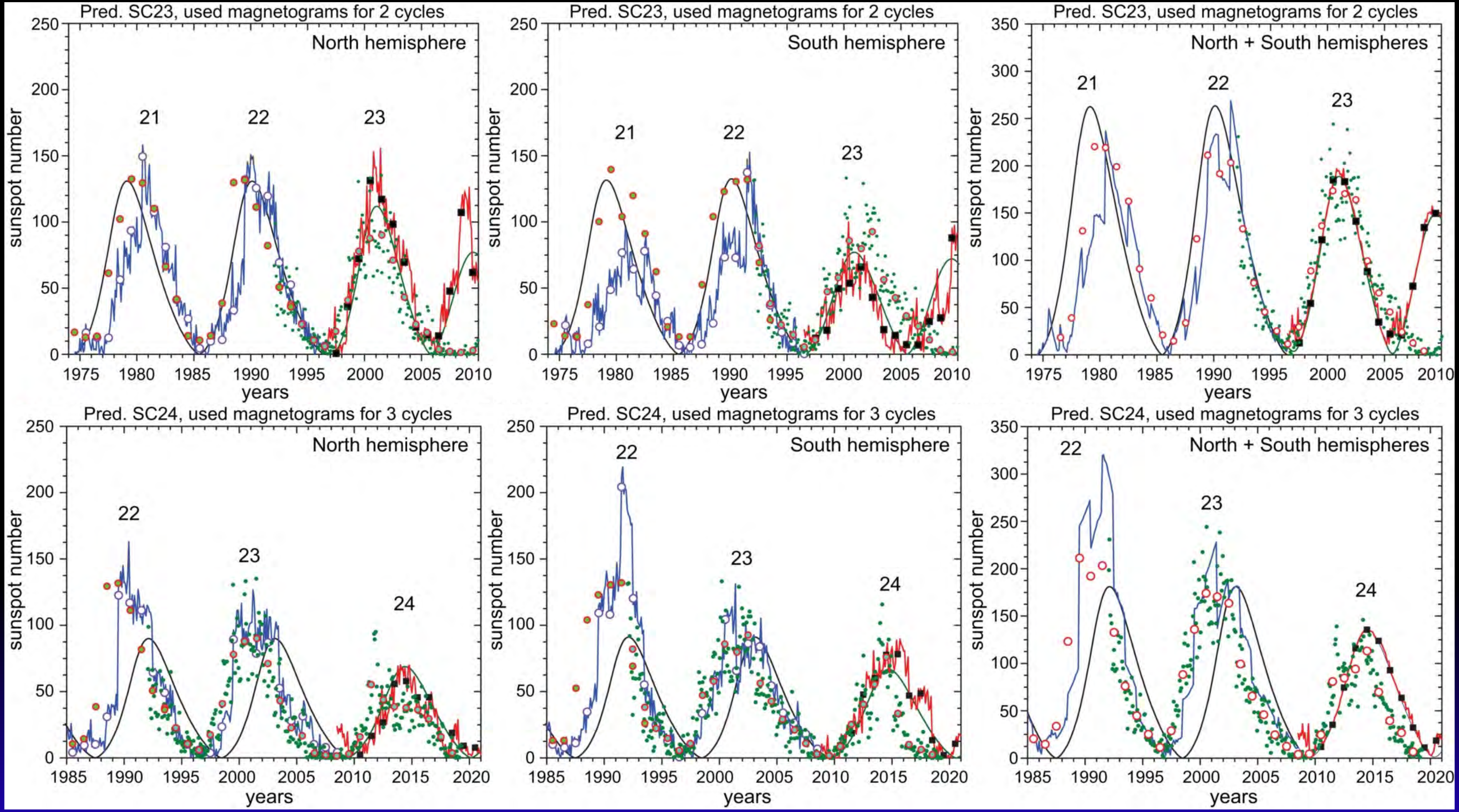
Toroidal and poloidal magnetic field components



Kitiashvili, 2020

‘Test’ predictions of Solar Cycles 23 and 24

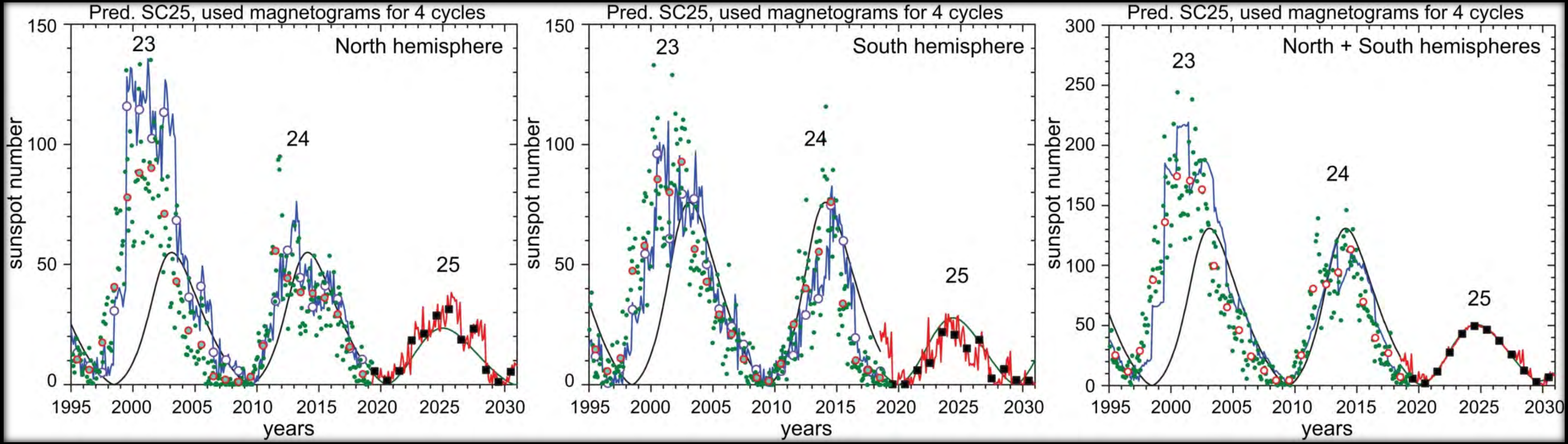
Data Assimilation



Kitiashvili, 2020

Solar Cycle 25

Data Assimilation



Solar Cycle 25 will be weaker than the current cycle and will start after an extended solar minimum during 2019 - 2021. The maximum of activity will occur in 2024 - 2025 with a sunspot number at the maximum of about 50 $\pm$ 15 with an error estimate of ~30%.

SC25 will start in the Southern hemisphere in 2020 and reach maximum in 2024 with a sunspot number of ~28 ($\pm$ 10%). Solar activity in the Northern hemisphere will be delayed for about 1 year (with error of $\pm$ 0.5 year) and reach maximum in 2025 with a sunspot number of ~23 $\pm$ 5 ($\pm$ 21).

Challenges of Solar Activity Forecasting

- Limited knowledge about past and current global solar activity
- Short time series of available observations
- No realistic theoretical description of global solar dynamics
- Evaluation quality of a forecast
- Estimation errors and uncertainties both in observations and models
- Physics- and observation-based global solar activity characterization (new standard development)

Work Plan to build reliable forecasts of Solar Activity

1) Continue synoptic observations of solar surface and subsurface dynamics

2) Development of new data analysis techniques
   a) procedures for error estimates and observational data reconstruction
   b) hybrid AI/DA methods
   c) methods of cross-analysis of different types of observations

3) Development of a methodology to extract information about magnetic field strength, distribution, and dynamics from helioseismic inferences

4) Development of global first-principle models of solar dynamics and dynamo