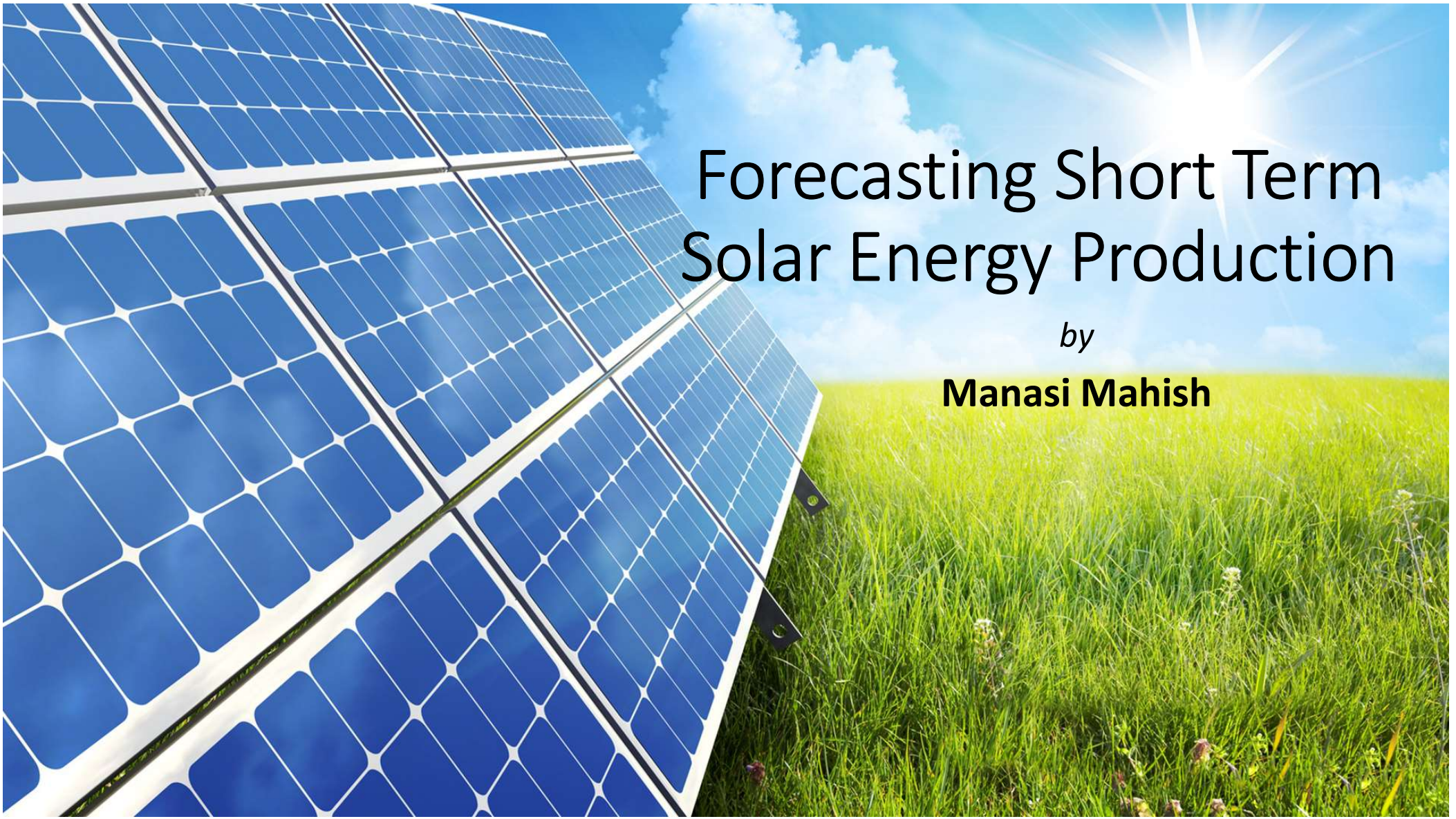


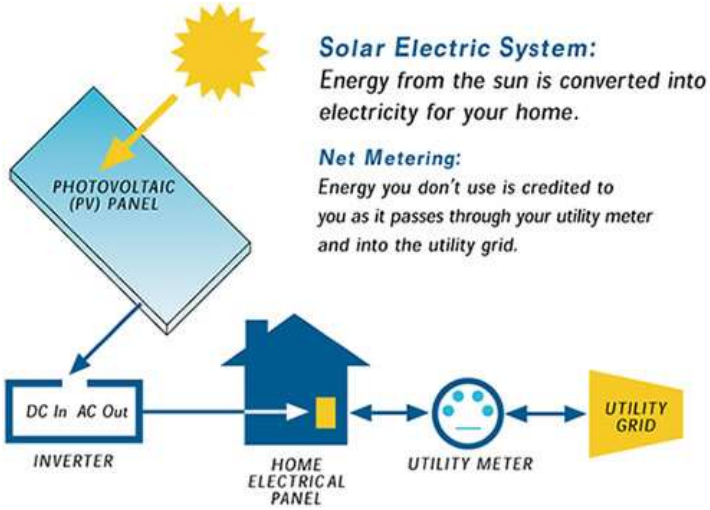
Forecasting Short Term Solar Energy Production

by

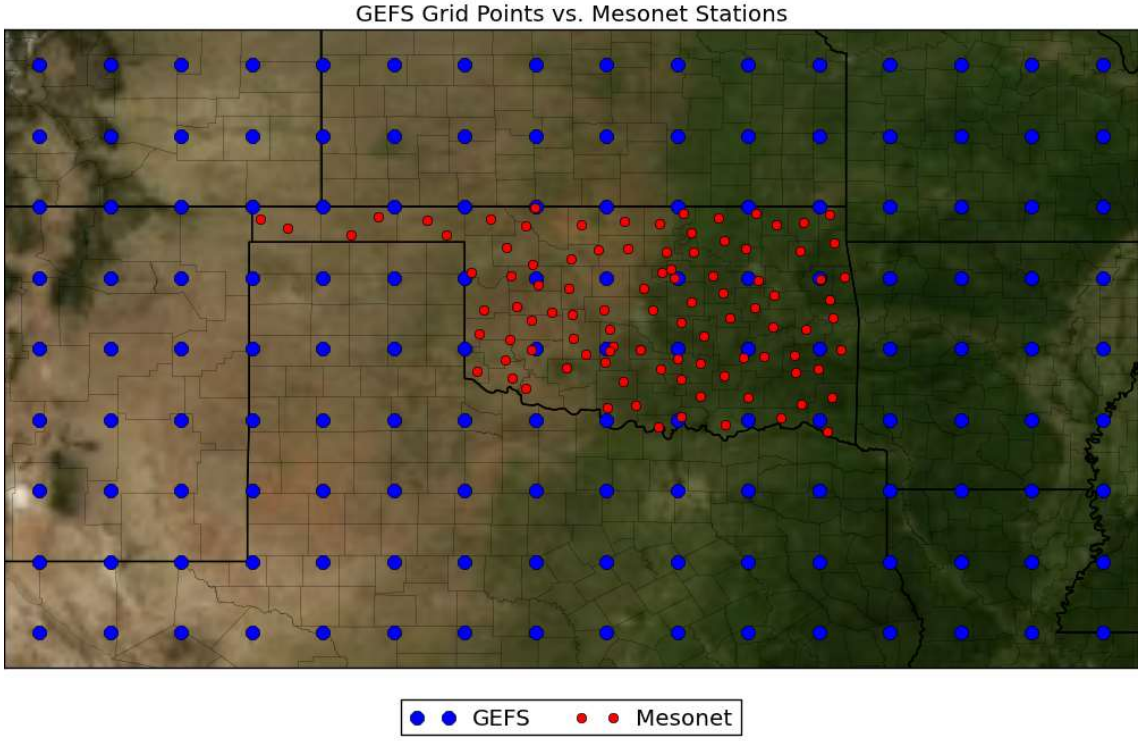
Manasi Mahish

Solar energy production: current situation

- Advantage
  - renewable
  - readily available
  - low carbon footprint
- Challenge
  - dependent on weather variables
  - difficult to accurately predict time ahead
  - may cause the utilities to purchase emergency power from neighboring utilities while over-predicting
  - may cause excess generation of power while under-predicting
- Solution
  - forecast solar energy production as accurate as possible using machine learning algorithms



Scope of the work

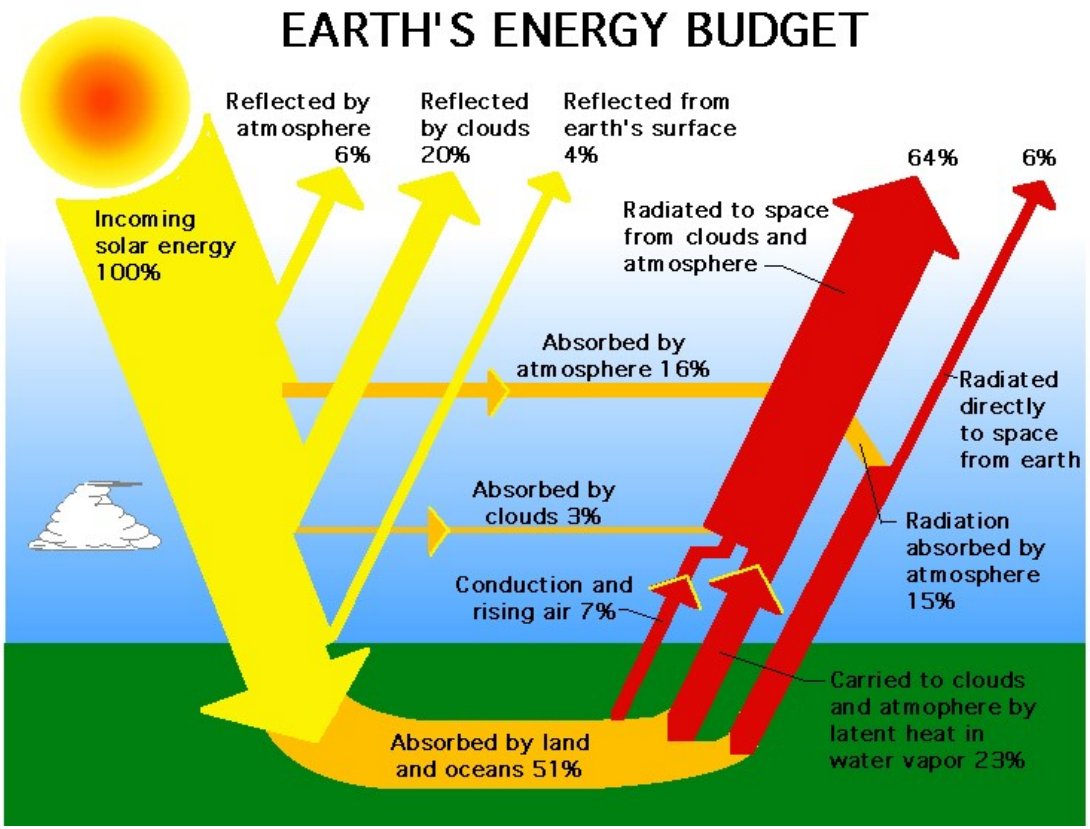


- Forecast solar energy production at 98 Mesonet stations (red)
- Model: machine learning algorithm- linear regression, gradient boosting, random forest
- Input data: Numerical weather prediction (NWP) model data from NOAA/ESRL Global Ensemble Forecasting System (GEFS) for 1° apart grid points (blue)
- Target metrics: Normalized root mean square error (NRMSE) as 0.05 and correlation coefficient as 0.9 between predicted and actual solar energy production

Data description

- Data structure: 5113 (days) x 11 (model ensemble) x 5 (times /day) x 9 (latitude) x 16 (longitude)
- Input data
  - Radiative flux: downward long-wave radiative flux at surface (DLWRFS), downward short-wave radiative flux at surface (DSWRFS), upward long-wave radiation at surface (ULWRFS), upward short-wave radiation at surface (USWRFS), upward long-wave radiation at top of the atmosphere (ULWRF)
  - Temperature: maximum temperature at 2 m above ground (MaxT), minimum temperature at 2 m above ground (MinT), current temperature at 2 m above ground, temperature of surface (T)
  - Other variables: air pressure at mean sea level (Pr), 3-Hour accumulated precipitation at surface (Precip), precipitable water over entire depth of atmosphere (PW), specific humidity at 2 m above ground (H), total cloud cover over entire depth of atmosphere (TCC), total column-integrated condensate over entire atmosphere (TC)
- Target variable
  - Solar energy production
- Training data: 1997 – 2007 weather (input) and solar energy (output) data
- Model prediction: 2008 – 2012 solar energy production based on weather data

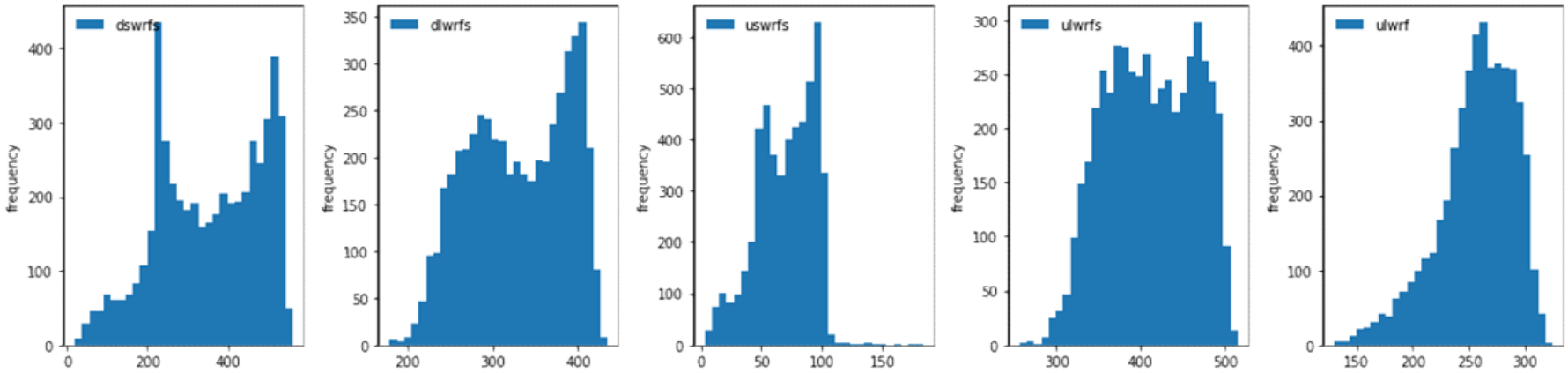
Exploratory analysis: radiative flux



https://kisialevelgeography.wordpress.com/as-atmosphere-weather/

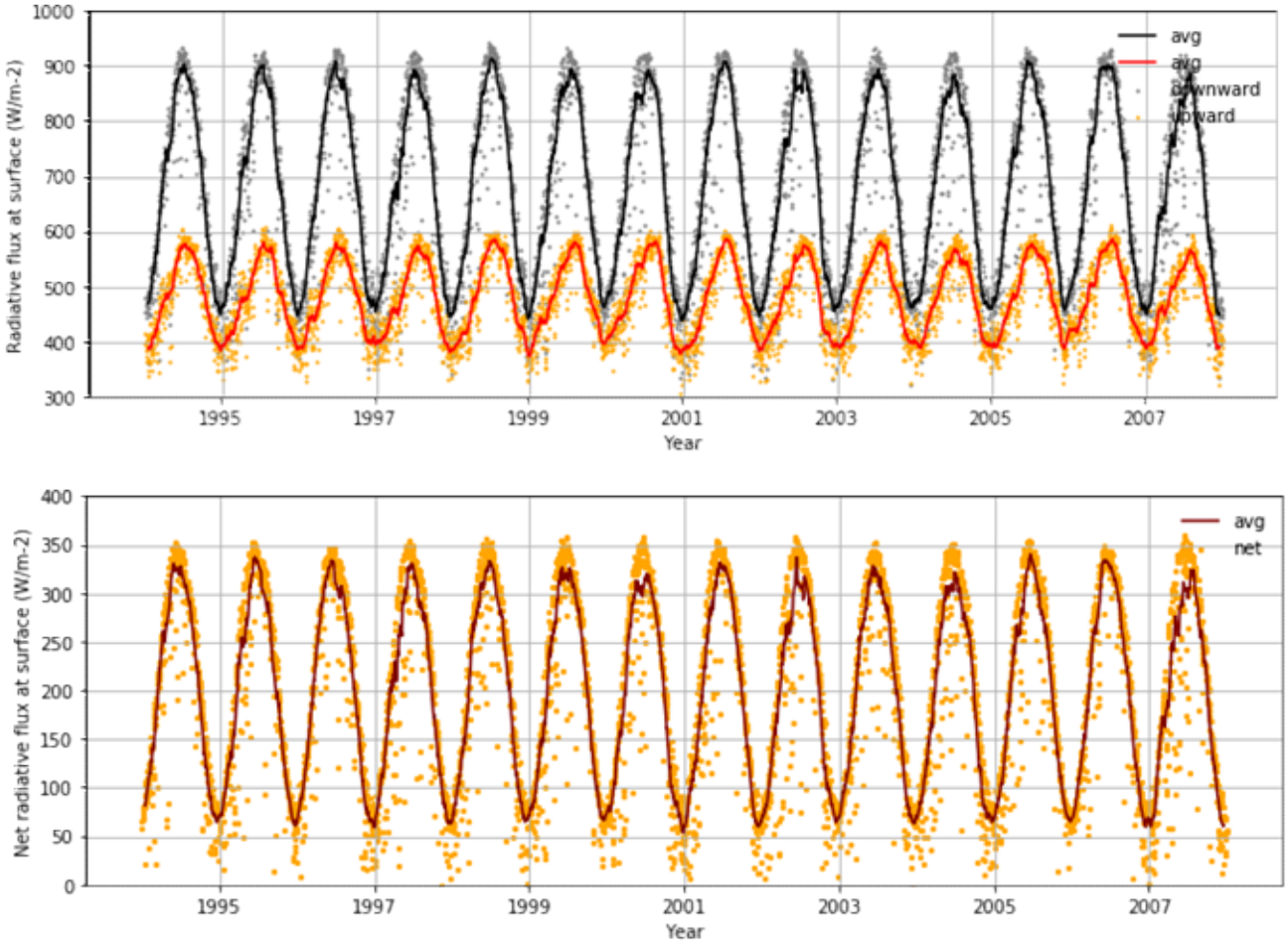
- Radiative Flux: Amount of solar energy radiated through a given area
- Incoming radiation ~ shortwave, high energy, emitted from very hot sun
- Upward radiation ~ longwave, low in energy, re-emitted by much cooler earth
- A small fraction goes back to space

Frequency distribution: radiative flux



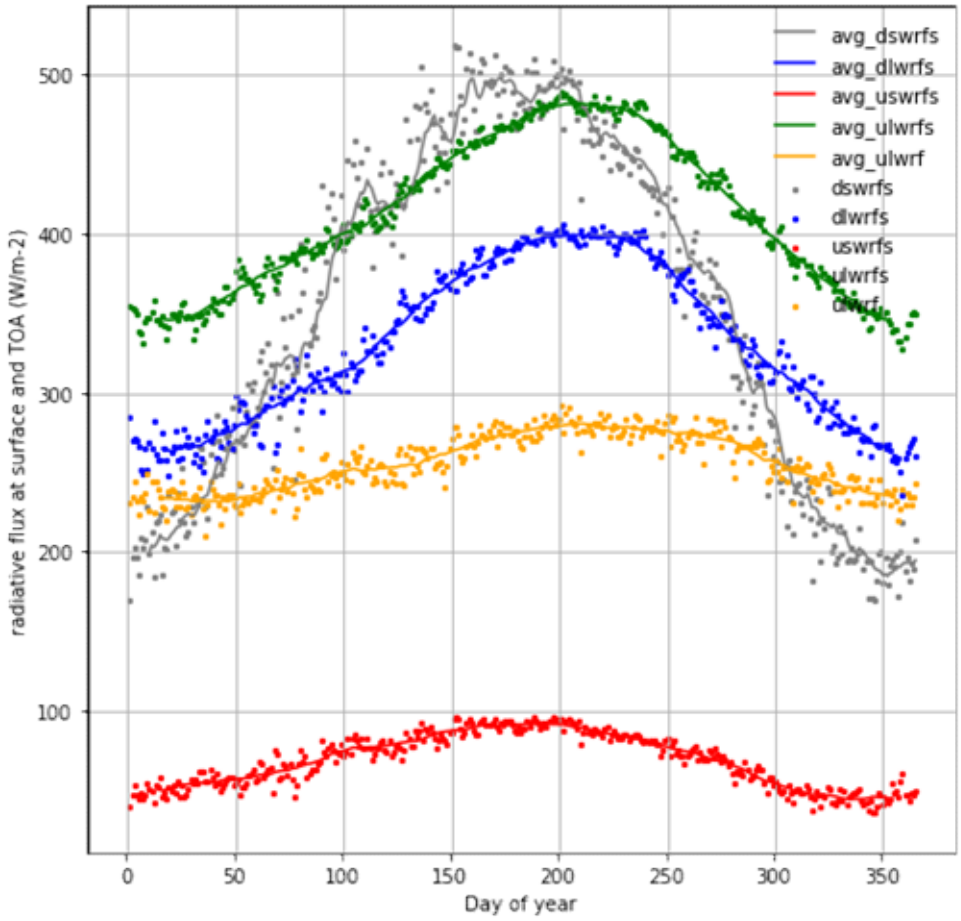
- Downward flux with 2 distinct peaks - summer and winter
- Downward shortwave - highest magnitude
- Upward shortwave– lowest magnitude
- Upward flux – longwave primarily

Time series analysis: radiative flux



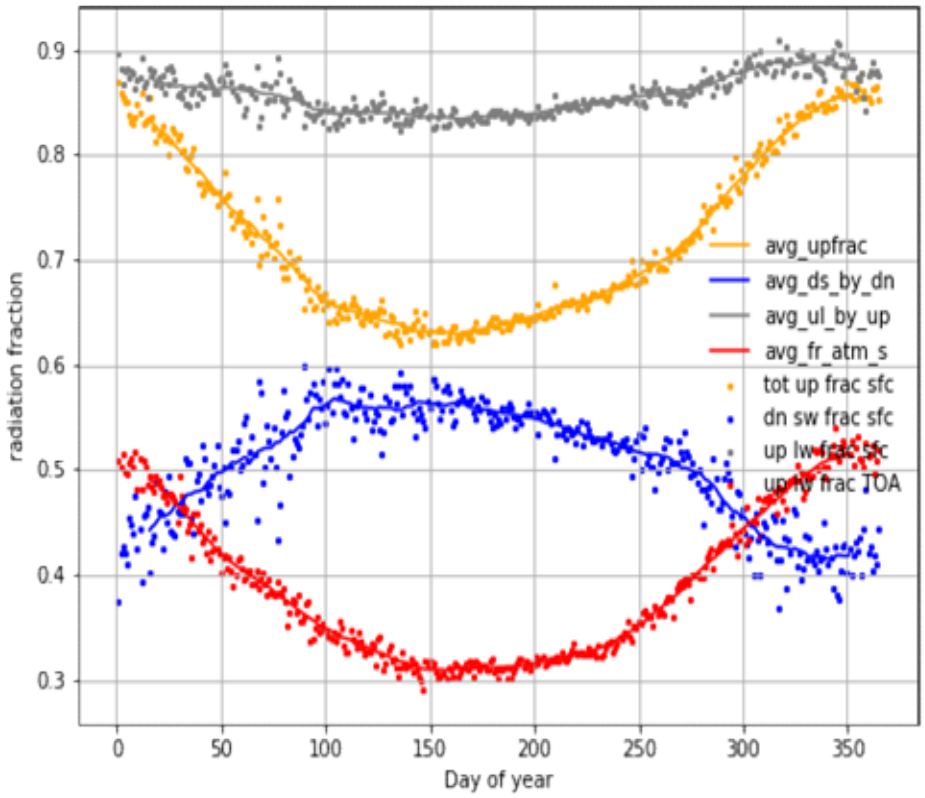
- Downward radiation: shortwave (mainly) + longwave
- Upward radiation: shortwave + longwave (mainly)
- Net radiation at surface: downward – upward
- Seasonality evident in downward, upward and net radiation at surface
- Seasonality caused by tilt of earth (primarily) + various weather elements

Seasonal variation in radiative flux



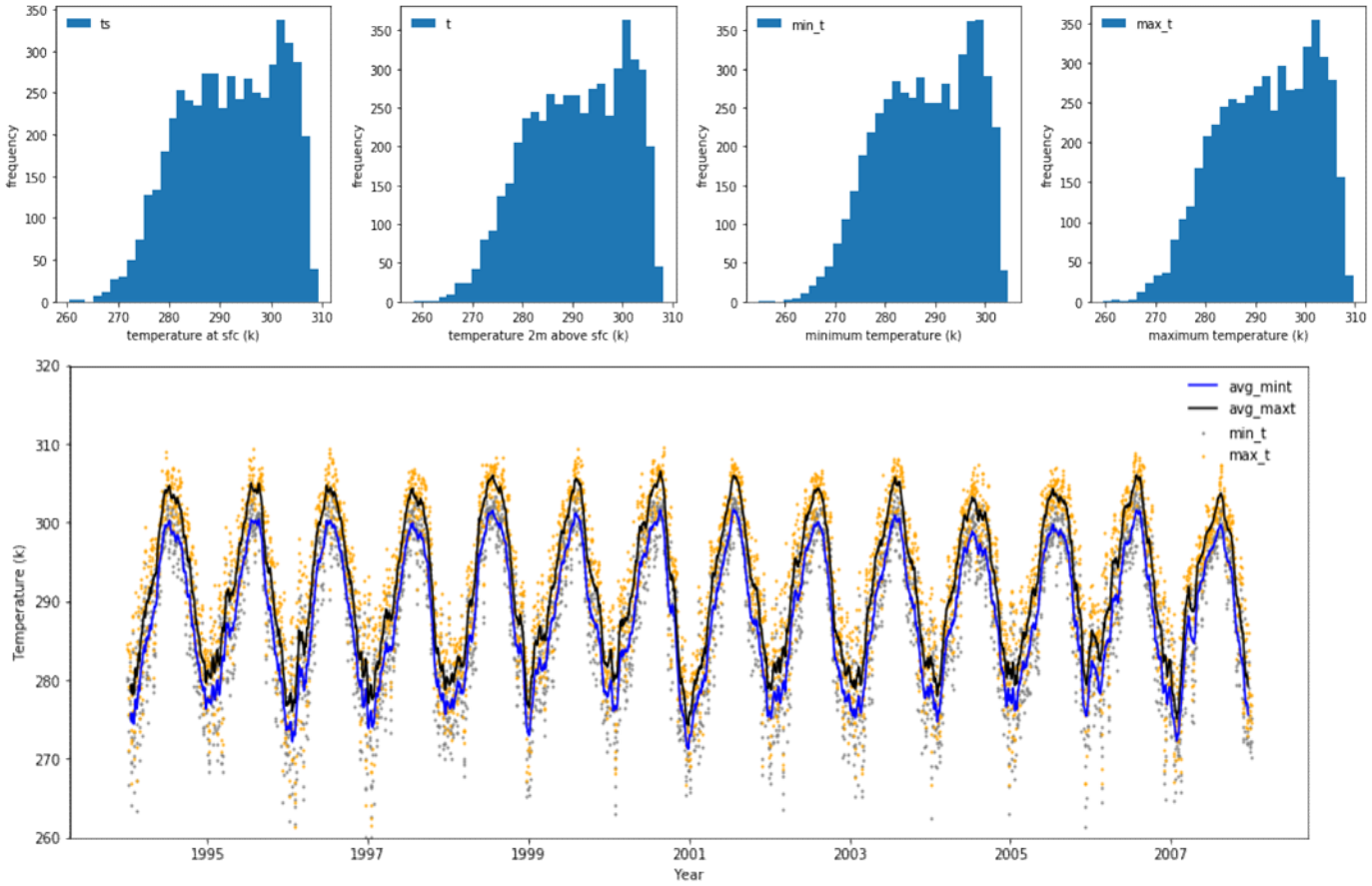
- Seasonality in downward shortwave is greatest
  - A change of 200 to 500 units between winter low to summer high
- All radiations high in summer and lower in winter
- There is time lag between the peaks of various radiations
  - Re-emitted long wave peaks later than the shortwave

Seasonal variation in radiation fraction



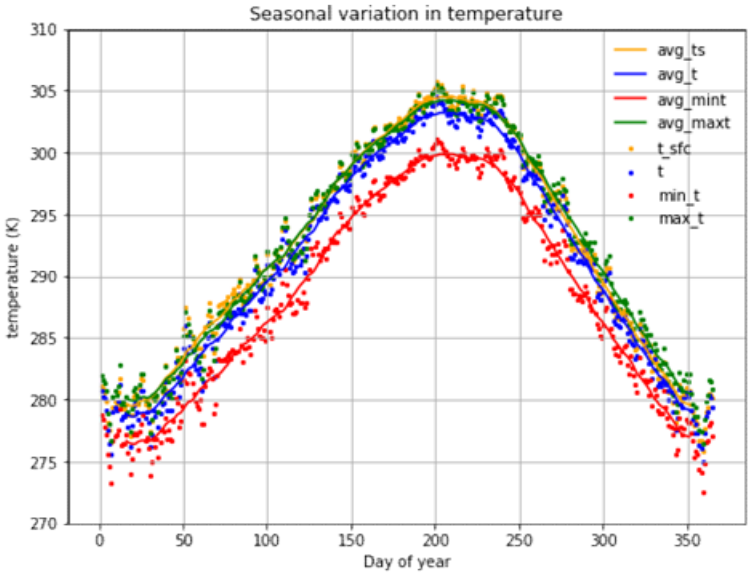
- The fraction of energy re-emitted from earth surface (orange) is less in summer
- Downward shortwave radiation as fraction of total downward radiation (blue) is higher in spring and summer
- Upward longwave radiation as fraction of total upward radiation (gray)> 80%
- Upward longwave radiation at TOA as a fraction of total downard radiation at surface (red) is less in summer months

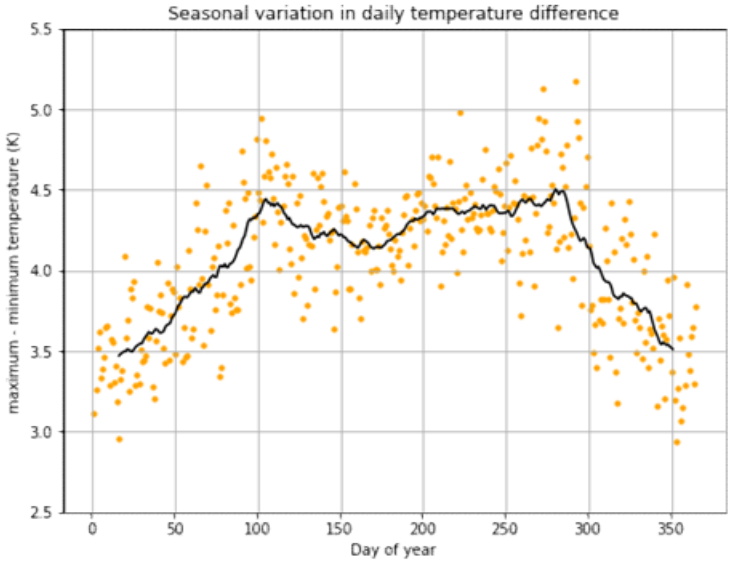
Frequency & time series analysis: temperature



- Distribution similar for all temperature measurements
- Temperature closely tied to the amount of radiation at surface
- Seasonality evident with high in summer and low in winter

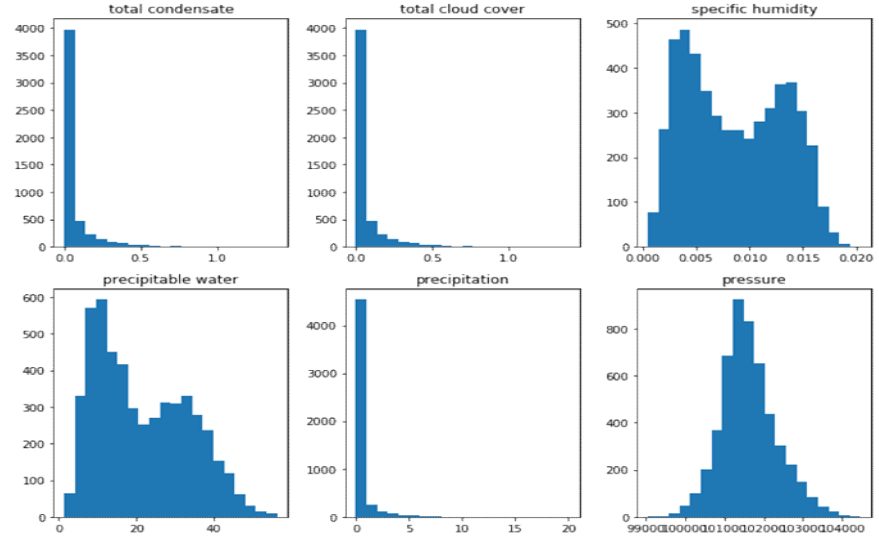
Seasonal variation in temperature

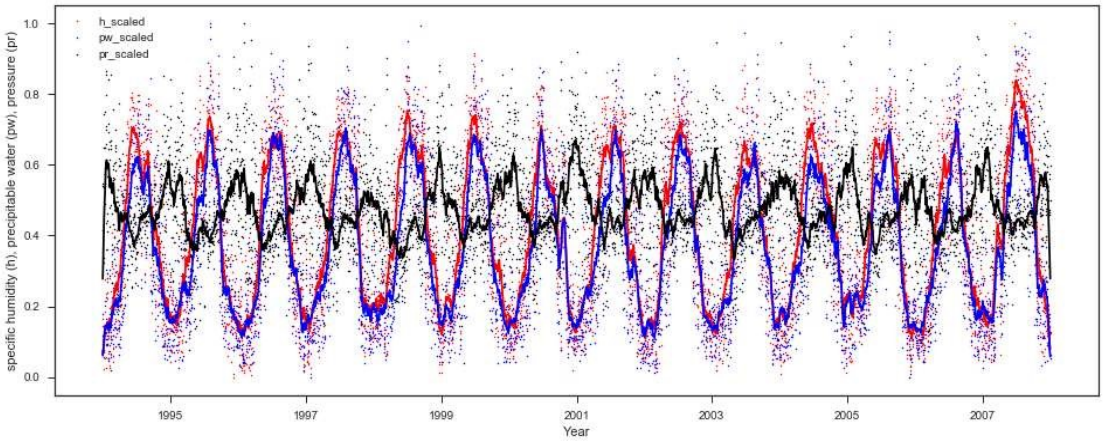




- Seasonality evident in temperature
- Temperature at surface (orange) higher than temperature above surface (blue) due to higher temperature gradient near surface
- Difference in daily maximum and minimum temperature higher during spring and fall, and lower in winter

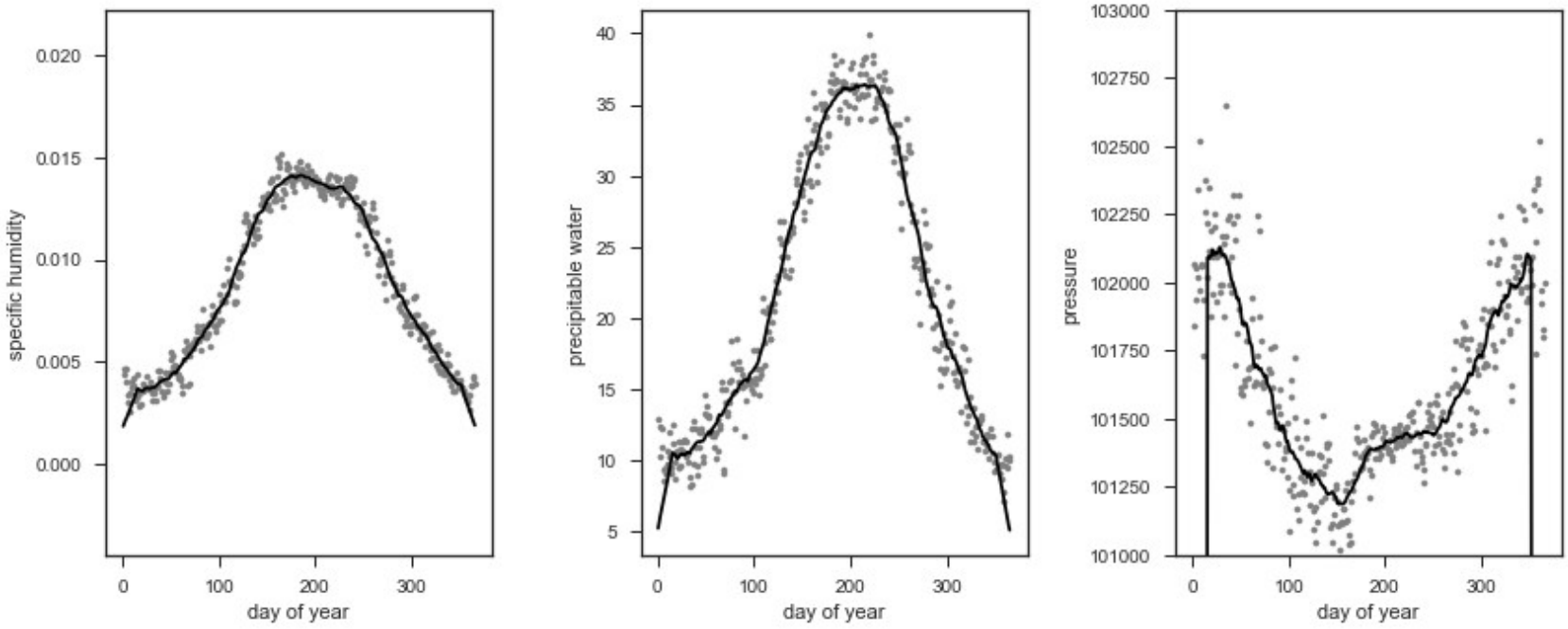
Frequency & time series analysis: other variables





- Total condensate, cloud cover and precipitation : most data near first bin
- Specific humidity and precipitable water with bimodal distribution, similar to temperature, large number of humid and dry days
- Pressure more normally distributed
- Seasonality in specific humidity, precipitable water and pressure
- Specific humidity and precipitable water high in summer low in winter
- Pressure high in winter and low in summer

Seasonal variation in humidity and pressure



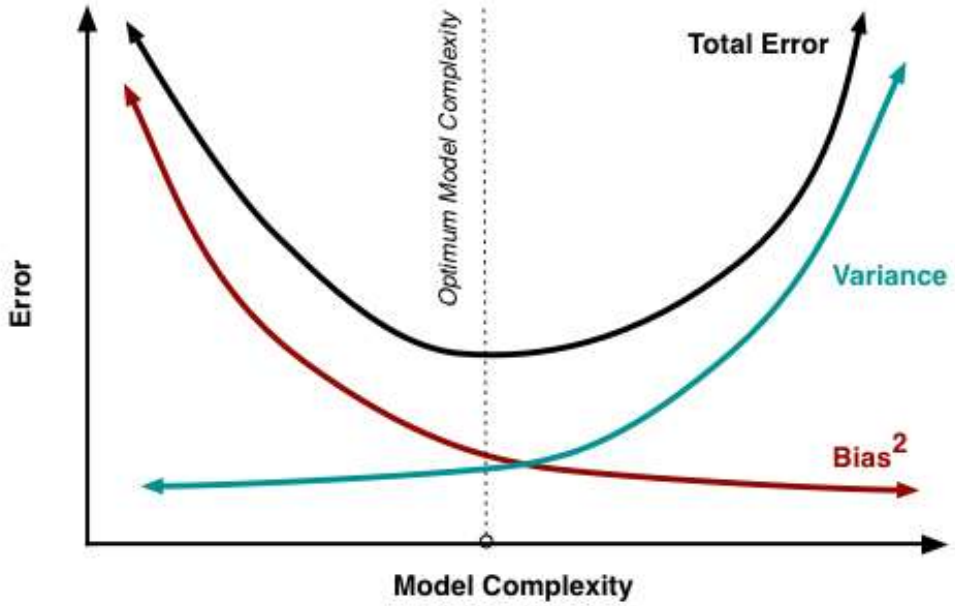
- Seasonality evident in specific humidity, precipitable water and pressure – specific humidity, and precipitable water high in summer and low in winter, pressure high in summer and low in winter

Correlation analysis with the target variable

Variable	Correlation coefficient with target	Variable	Correlation coefficient with target
Target	1.000	H	0.432
DSWRFS	0.888	DLWRFS	0.351
USWRFS	0.843	PW	0.254
ULWRF	0.660	Pr	-0.227
TS	0.644	Precip	-0.309
ULWRFS	0.633	TC	-0.487
MaxT	0.628	TCC	-0.487
T	0.627	MinT	0.580

- The radiations in general strongly correlated with solar energy
- Downward shortwave radiation primary source of energy - highest correlation with solar energy production
- Pressure having minimum / no effect on solar energy
- Total cloud cover (TCC), total condensate (TC), precipitation negatively affect solar energy; more atmospheric moisture, more absorption of radiation by the atmosphere, less energy reaching surface

Model prediction: choosing the right model



- Simple model
  - Underfitted dataset
  - High bias
  - Low variance
  - Generalized
- Complex model
  - Overfitted dataset
  - Low bias
  - High variance
  - Not generalized
  - Flexible
- Objective: build a model with optimum model complexity

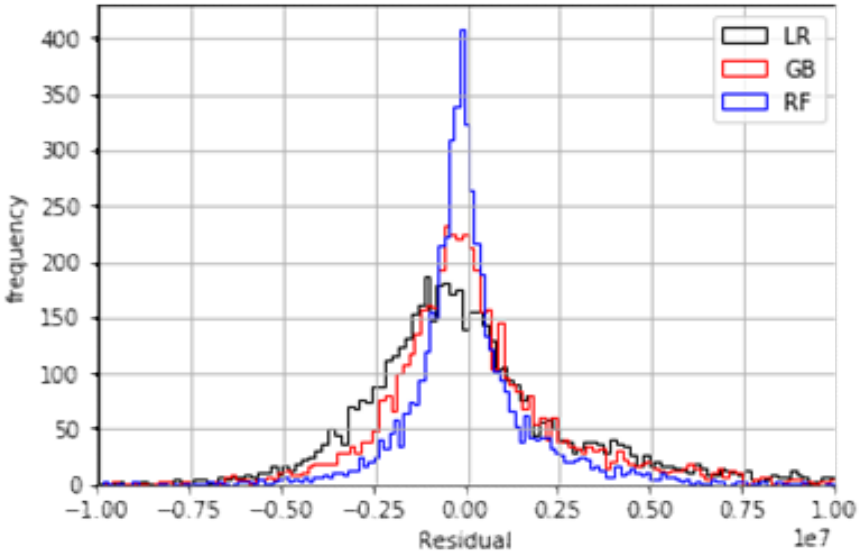
Machine learning algorithm

- Linear regression
  - Simple model
  - Assumed linear relations between features and target variable
  - Bias high, variance low
- Gradient boosting
  - Built on weak regressors
  - Sequential process
  - Following regressor learns from previous regressor’s mistake
- Random forest
  - Decision trees prone to overfitting
  - Parallel process
  - Higher number of decision trees employed to process re-sampled data
  - Relies on voting to average out overfitting
  - Computationally expensive

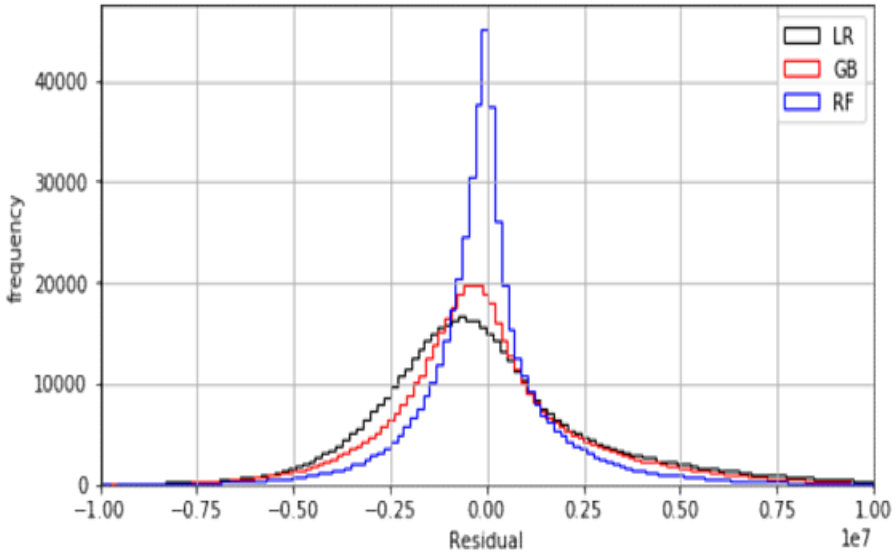
Modeling steps

- Train (70%), test (30%) split in input dataset
- Standardize data
- Gradient boosting: 'n_estimators': 10, 'max_depth': 5, and 'learning_rate': 0.4
- Random forest: ‘n_estimator’: 15
- Fit the training dataset and test model accuracy with test dataset
- The performance metrics : NRMSE and r
- Normalized root mean square error (NRMSE) between predicted and measured target variables
- $NRMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{pred.i} - y_{test.i})^2}{n}} \Big/ \frac{\sum_{i=1}^{n} {y_{test.i}}^2}{n}$
- Pearson correlation coefficient (r) between predicted and measured target variables
- $r = \frac{n\sum_{i=1}^{n} y_{pred.i}\ y_{test.i} - \sum_{i=1}^{n} y_{pred.i} \sum_{i=1}^{n} y_{tets.i}}{\sqrt{\left(n\sum_{i=1}^{n} y_{pred.i}^2 - \left(\sum_{i=1}^{n} y_{pred.i}\right)^2\right)\left(n\sum_{i=1}^{n} y_{test.i}^2 - \left(\sum_{i=1}^{n} y_{test.i}\right)^2\right)}}$
- Predict using unseen input weather dataset using the trained model

Residual distribution of ML algorithms



ACME station



All 98 stations

Comparisons of ML algorithms

Machine learning	Error components		Parameters	NRMSE / Correlation coefficient					
				Station 'ACME'				All stations	
	bias	variance		All features		Radiative flux only		All features	
				NRMSE	Corrcoef	NRMSE	Corrcoef	NRMSE	Corrcoef
Linear regression	high	low	default	0.191	0.912	0.2	0.901	0.2	0.909
Gradient boosting	high	low	n_estimator = 10	0.194	0.934	0.194	0.902	0.204	0.932
Random forest	low	high	n_estimator = 15	0.19	0.967	0.201	0.958	0.206	0.961

Conclusion

- The target NRMSE (0.05) was not achieved but the target correlation coefficient (0.9) was obtained in all cases
- Solar energy was predicted appropriately using weather inputs
- Radiative data alone effectively predicts solar energy with some error margin
- Similar accuracy achieved using 3 ML algorithms
- Linear regression is the simplest and chosen model for solar energy prediction


Thank
you