CDAF Task Team

Jonathan Mittaz

Prasanjit Dash, Rishi Gangwar,

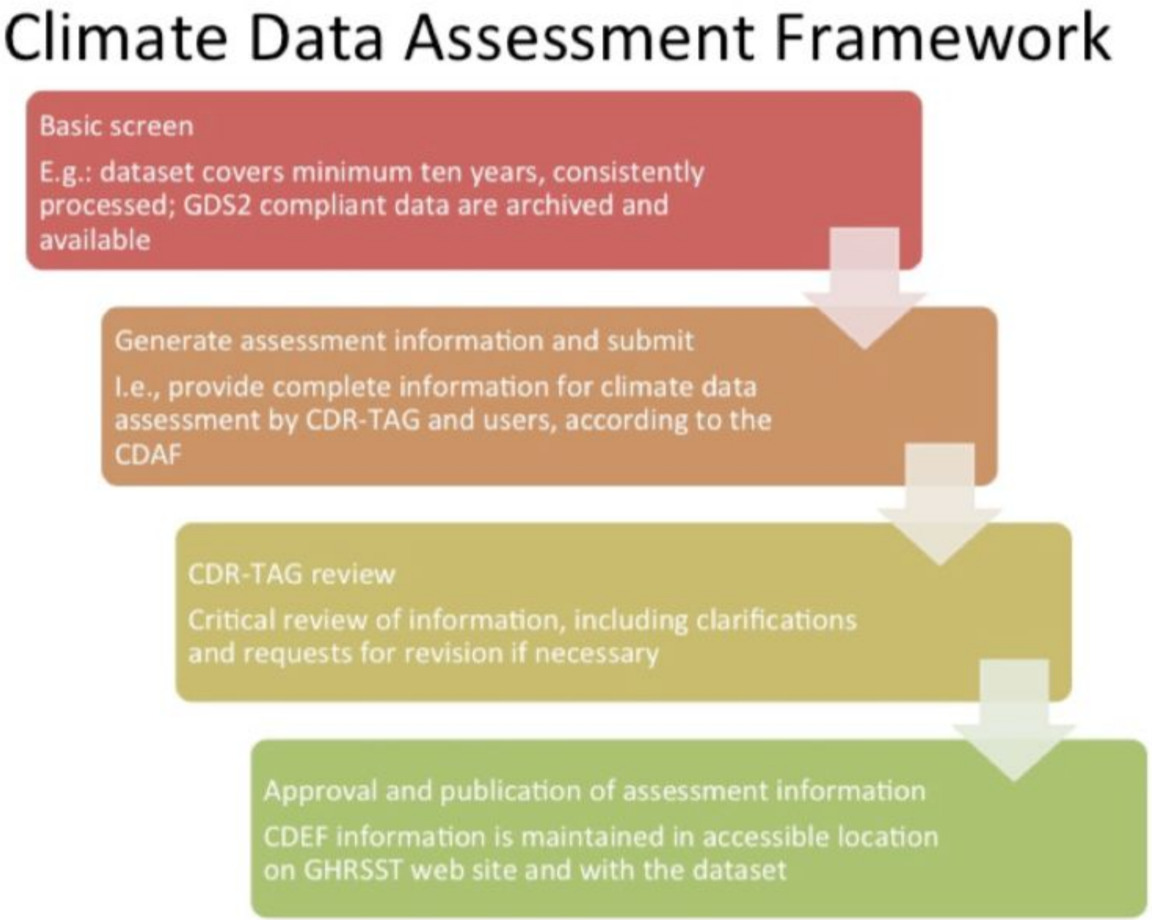
Jean-Francois Piolle, Ed Armstrong

The CDAF

- Climate Data Assessment Framework
  - A CDR is "a time series of measurements of sufficient length, consistency and continuity to determine climate variability and change" (NRC, 2004), ideally traceable to SI standards.
  - The CDAF is intended to be used to support users of sea surface temperature (SST) datasets to understand the suitability of GHRSST datasets for use as Climate Data Records (CDRs).
    - Will provide authoritative, comparable information about GHRSST datasets that will allow users to make their own judgment about the use of the datasets as CDRs for their application.
- Task Team
  - Has been in existence for several years but other pressures/projects have meant that progress has been slow to date
  - Progress has been made this year on the coding side
    - Last report discussed MMD work which has been on hold for the past year

Overview of the CDAF

- There are a number of sources of information that are needed



CDAF Tool metrics

- Primarily to be used to provide information for section 2
  - Generate assessment information (Quantitative measures)
    - Systematic differences referenced to drifting buoys
      - *Global median difference*
      - *Bin SST to buoy differences on a 10^ox10^o scale and estimate standard deviation of subsets (filter out bins with too few matches first)*
    - Systematic differences referenced to Argo measurements
      - *Global median difference*
      - *Bin SST to buoy differences on a 20^o(lat)x90^o(lon) scale and estimate standard deviation of subsets (filter out bins with too few matches first)*
    - Non-Systematic effects
      - *Calculate robust standard deviations after median values (including geographic variations) have been removed using above statistics*
    - Stability
      - Use the GTMBA using simplified method from Merchant et al. (2012)
    - SST Sensitivity
      - Calculated SST sensitivities based on retrieval algorithm characteristics (provided by data provider)

Example of current output form

KEY DESCRIPTIVE FEATURES	INFORMATION
Period covered	
Geographic range	
Spatial resolution	
Temporal resolution	
Timeliness of new data	
Dataset volume	
Valid data fraction	
Data level / grid	
Observation technology	
Dependence on other data	
Type(s) of SST	
Traceability	
Uncertainty info in product	

QUANTITATIVE MEASURES	VALUE	COMMENTS
Difference relative to drifting buoys		*Global median difference of satellite minus drifting buoy SST, across full dataset. The satellite SSTs are SST_{skin} with no skin-effect adjustment, so a skin-effect difference of order -0.2 K is to be expected.*
Difference relative to Argo		*Global median difference of satellite minus upper Argo float SST, across full dataset. The satellite SSTs are SST_{skin} with no skin-effect adjustment, so a skin-effect difference of order -0.2 K is to be expected.*
Geographical variation in difference relative to drifting buoys		*Geographical variation in difference, as described by the standard deviation of median satellite minus drifting buoy SST differences on space scales of ~1000 km, across the full dataset.*
Geographical variation in difference relative to Argo measurements		*Geographical variation in bias, as described by the standard deviation of median satellite minus upper Argo float SST differences on space scales of 20° latitude by 90° longitude, across the full dataset.*
Dispersion relative to drifting buoys		*Spread of differences associated with non-systematic effects as quantified by a robust standard deviation of differences of satellite and drifting buoy data, after removing the geographical variations in differences quantified above*
Stability		*95% confidence interval for the relative multi-year trend between satellite SSTs and the Global Tropical Moored Buoy Array*
Sensitivity to true SST		*Average weight of the satellite observations in determining SSTs in the dataset, the difference from 100% representing the weight of prior information in the SSTs*

AVAILABILITY, DOC'N, FEEDBACK	
Data URL / ftp / DOI	
Primary peer reviewed reference	
Source of technical documents	
Dataset restrictions	
Facility for user feedback	
Other documentation	

OTHER PRINCIPLES (GCOS)	COMMENTS
2. and 12. Overlaps between sensors exist and are exploited to harmonize the dataset	
3. Detailed history of methods/ algorithms is available	
11. Constant sampling within diurnal cycle	

Workplan

- **Get reference data**
- **Produce Matchups**

Together with Matchup Task Team (Jean-Francois)
Matchup production coordinate with Matchup TT
(Felyx based)
Initial data produced but not integrated

- **Create Statistics**

As much as possible this will use already existing code/processes.

- Currently have rewritten (Python) versions of some of Gary Corlett's IDL code, work ongoing

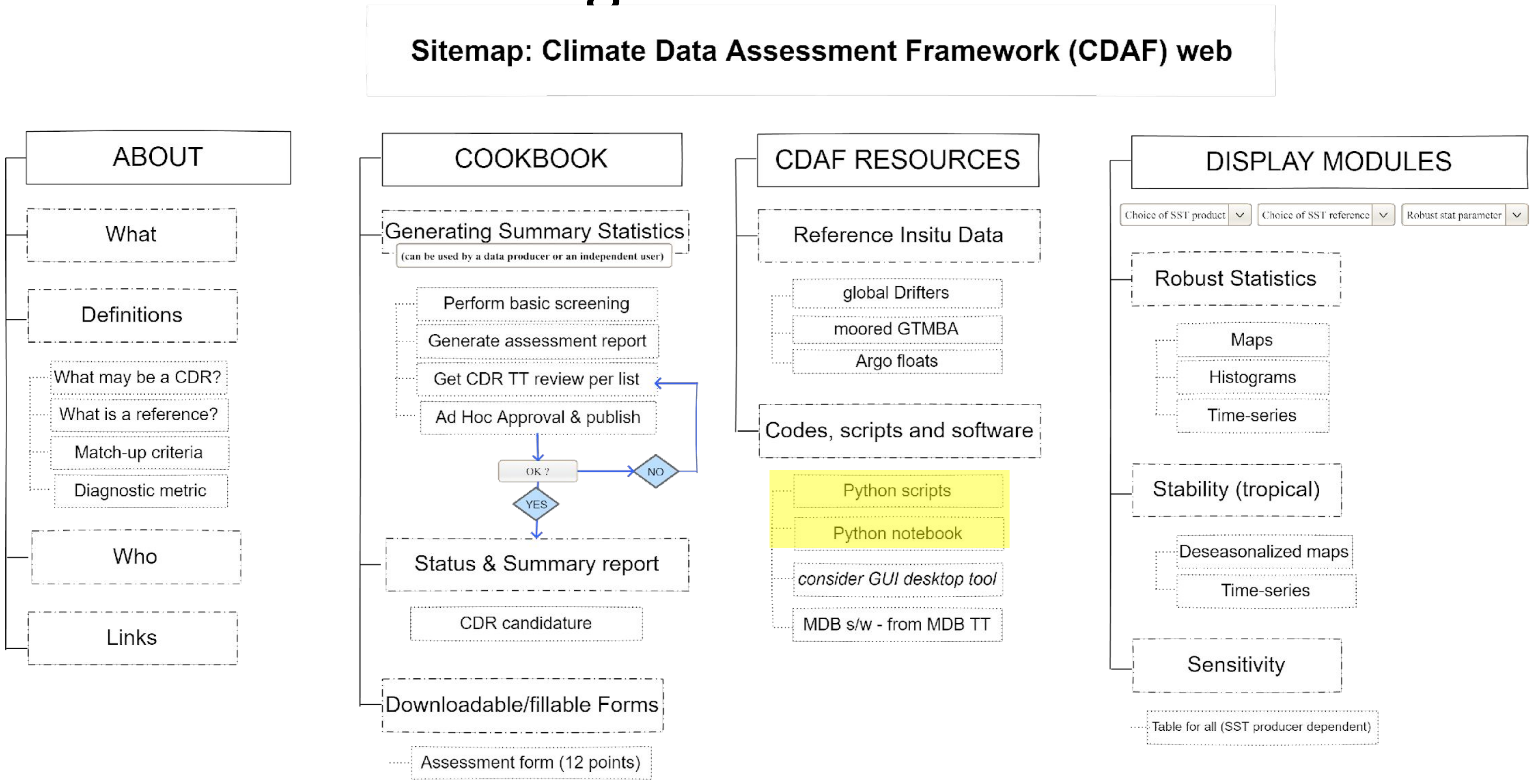
The current set of metrics are fairly simple

- Means/Medians
- Standard Deviation/Robust Standard Deviations
- Geographic binning code

- **Web/Form output**

Initial designs ideas created
(Prasanjit)

Web Workflow Design



IDL to Python Conversion

IDL codes' names (Gary C.)	Python Notebooks (new)
Main program cci_val_plot_master_timeseries_cdaf.pro subroutines/functions cci_cdaf_assessment_l4.pro rsd.pro stratify_data.pro calc_histogram.pro mpfitfun.pro cci_cdaf_assessment.pro	cci_val_plot_master_timeseries_cdaf.ipynb
Main program cci_stability_plot_main.pro subroutines/functions cci_cdaf_fit_trend.pro	cci_stability_plot_main.ipynb

In scripts and will also be available as Python notebooks
Extra statistics beyond original code available

Snippets

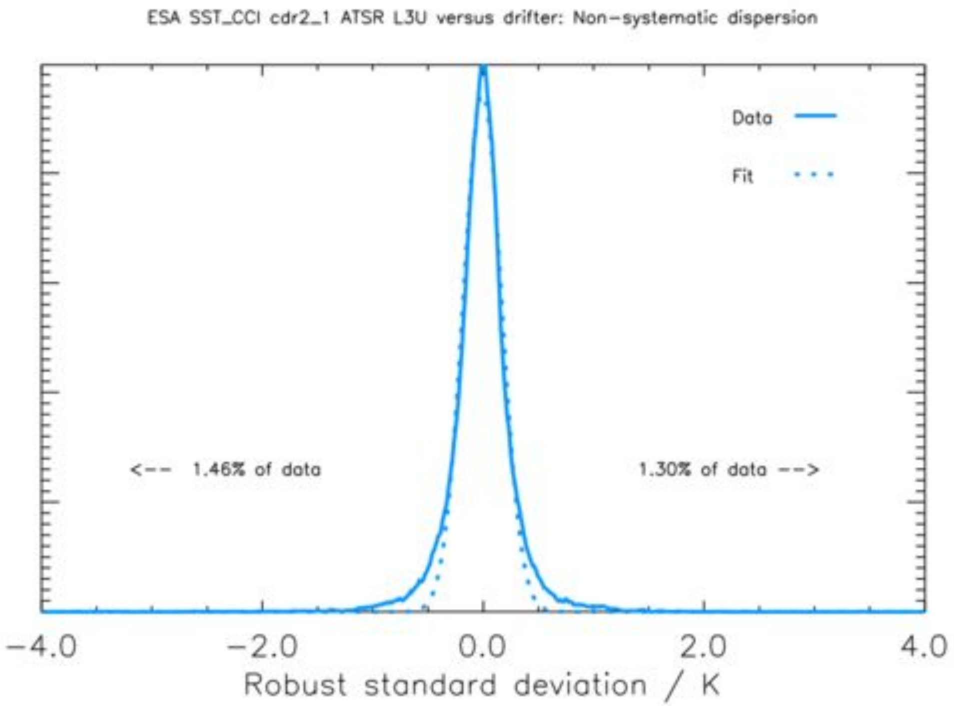
IDL code

```
function cci_cdaf_fit_trend, x, y, range, plot=plot, thick=thick, col=col

  COMPILE_OPT idl2

  years  = ((1./12.)*INDGEN(480)) + 1978

  id =  where(years ge range[0] and years le range[1] and finite(x) and finite(y), c)
  result = LINFIT(x[id], y[id], sigma = sigma)

  if keyword_Set(plot) then begin

    fit = result[0] + x[id]*result[1]
    oplot, x[id], fit, color=col, thick=thick
    oplot, x[id], fit+(1.96*sigma[1]*12), color=col, thick=thick, linestyle=1
    oplot, x[id], fit-(1.96*sigma[1]*12), color=col, thick=thick, linestyle=1
  endif

  return, [result[1] - 1.96*sigma[1], result[1]+1.96*sigma[1]]*12

END
```

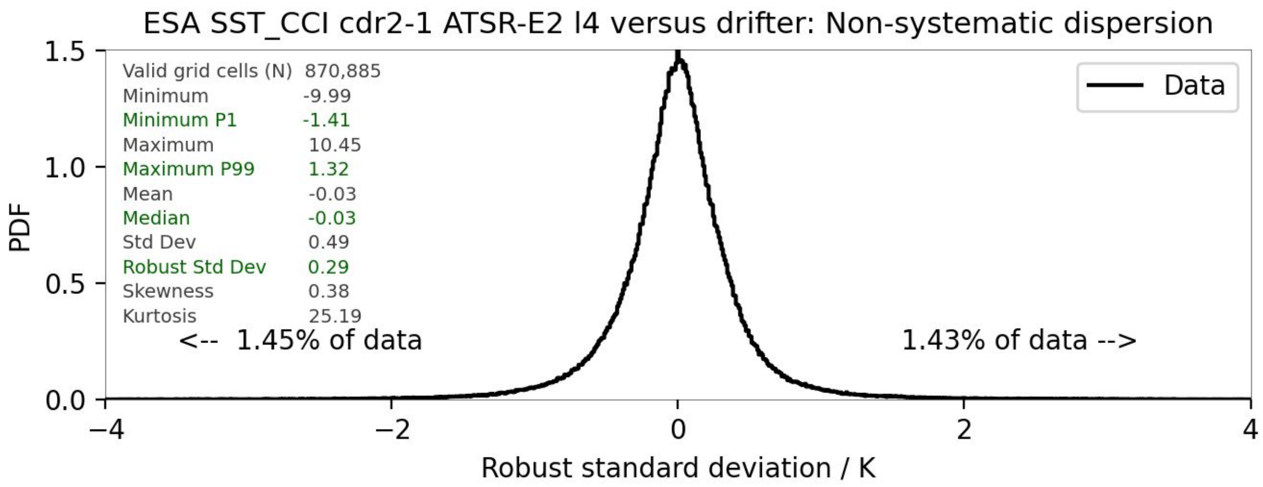
Python code

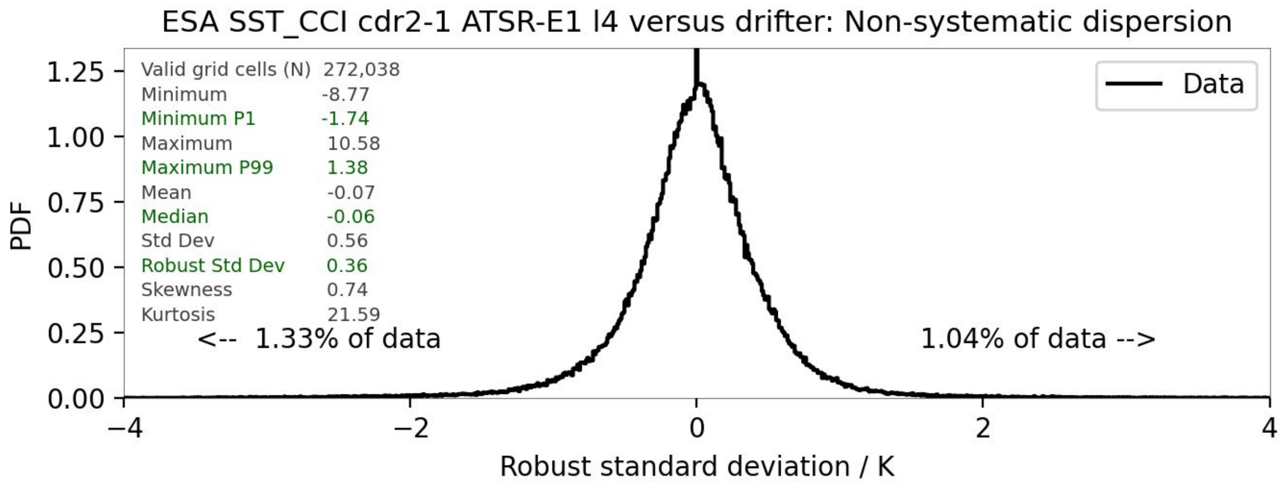
```python
def cci_cdaf_fit_trend(x, y, x_range, plot=False, thick=1, col='b', axes=None):
    years = (1.0 / 12.0) * np.arange(480) + 1978
    valid_indices = np.where((years >= x_range[0]) & (years <= x_range[1]) & np.isfinite(x) & np.isfinite(y))
    x_valid = x[valid_indices]
    y_valid = y[valid_indices]
    result = linregress(x_valid, y_valid)

    if plot:
        fit = result.intercept + x_valid * result.slope
        if axes is None:
            plt.plot(x_valid, fit, color=col, linestyle='--', linewidth=thick)
        else:
            axes.plot(x_valid, fit, color=col, linestyle='--', linewidth=thick)

        upper_bound = fit + 1.96 * result.stderr * 12
        lower_bound = fit - 1.96 * result.stderr * 12

        if axes is None:
            plt.fill_between(x_valid, upper_bound, lower_bound, color=col, alpha=0.2)
        else:
            axes.fill_between(x_valid, upper_bound, lower_bound, color=col, alpha=0.2)

    trend_range = [12 * (result.slope - 1.96 * result.stderr), 12 * (result.slope + 1.96 * result.stderr)]
    return trend_range
```

From Gary's IDL code



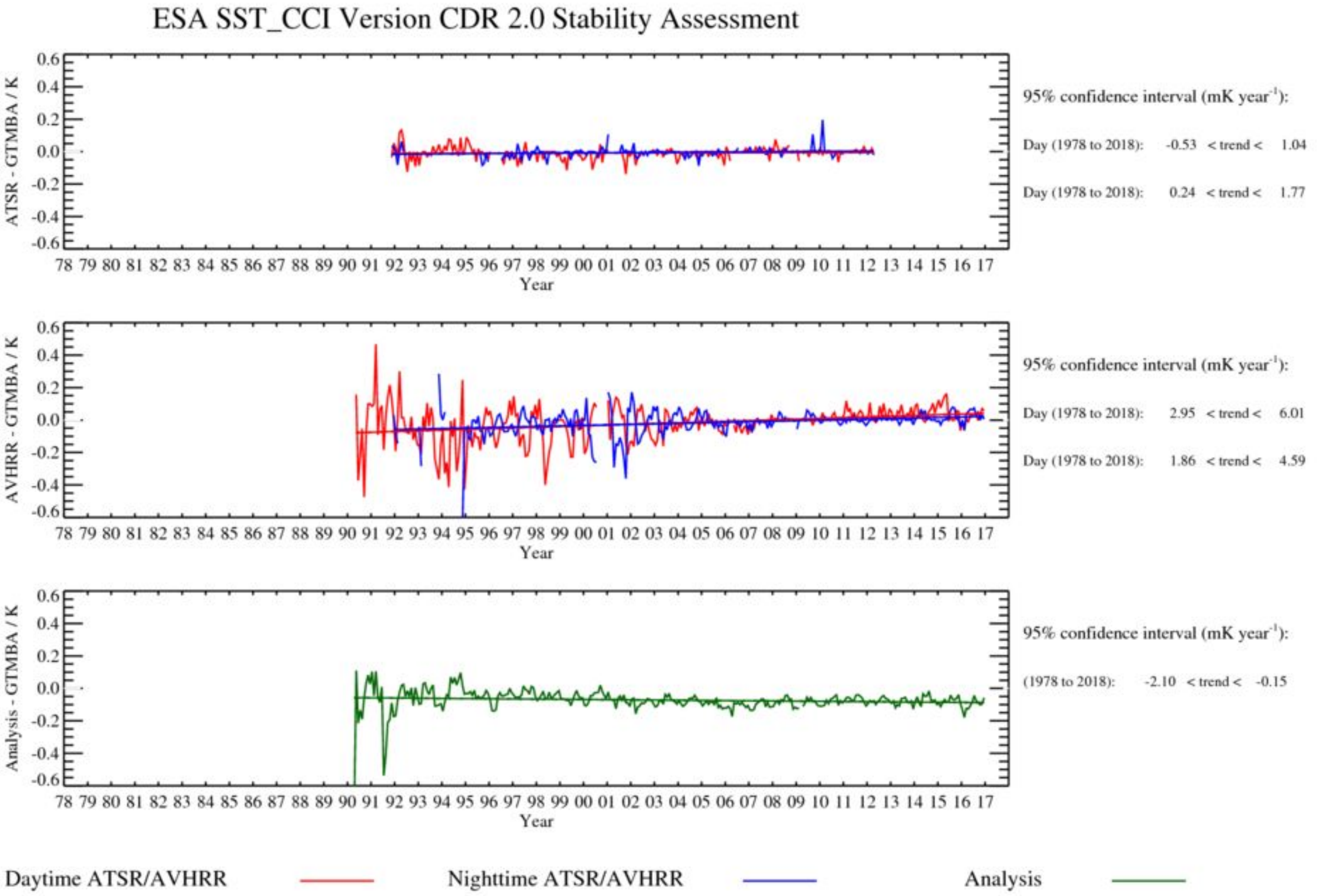
From Python code (newer)

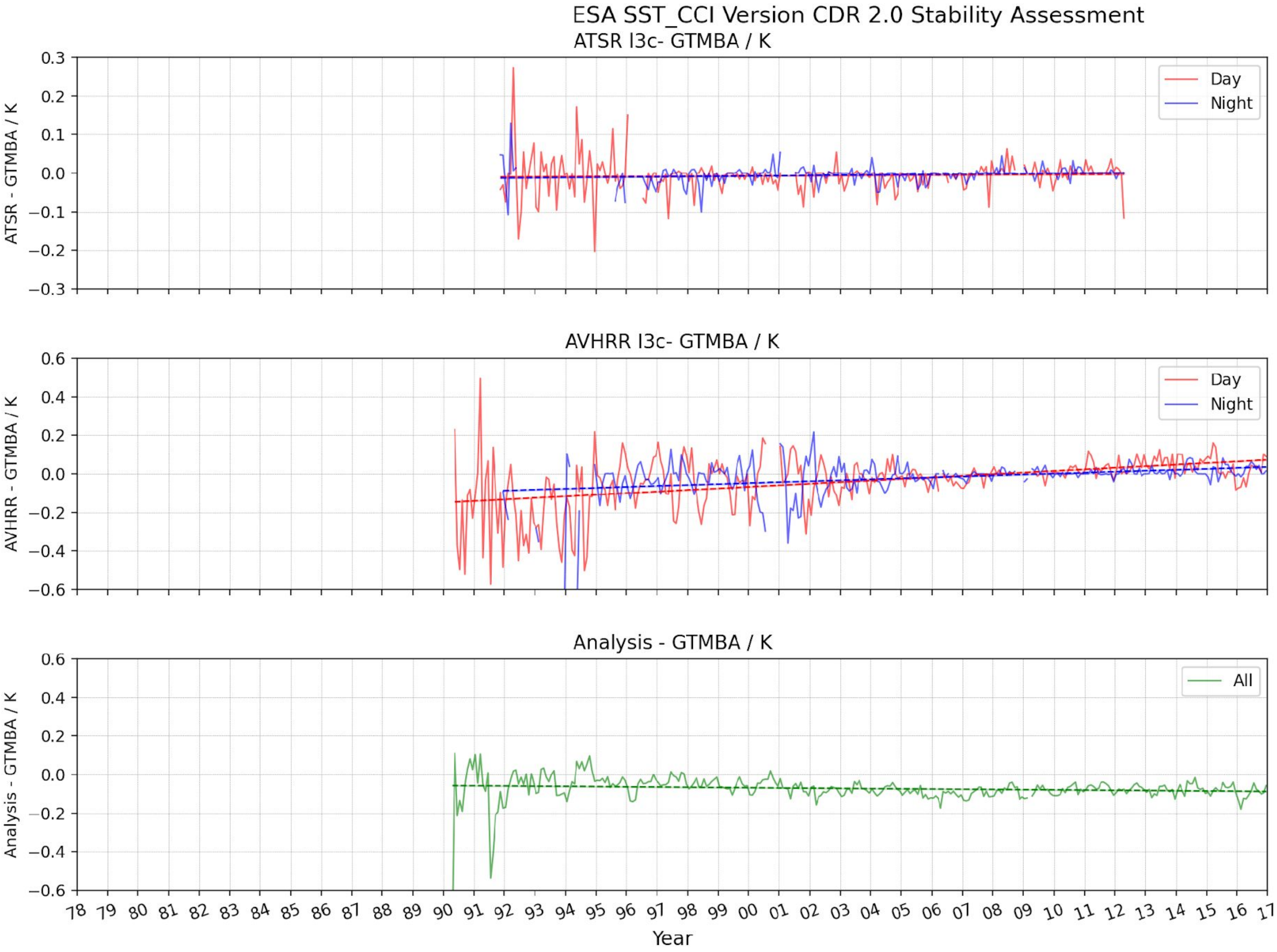




The parameters in Green are robust parameters

From Gary's IDL code





From Python code (newer)

95% confidence interval (mK year^{-1}):

Day (1978 to 2018): -0.61 < trend < 1.28

Night (1978 to 2018): 0.07 < trend < 1.32

95% confidence interval (mK year^{-1}):

Day (1978 to 2018): 6.49 < trend < 9.93

Night (1978 to 2018): 3.27 < trend < 6.51

95% confidence interval (mK year^{-1}):

(1978 to 2018): -2.10 < trend < -0.15

Next steps

- Remaining IDL2Py conversion as needed (12 months)

- Interface new code with new matchup data including extra statistics
  - Interface with Felyx generated MMDs
  - Analyze geo-spatial variations
  - Stability analysis

- Setup initial interface
  - Design already in progress
  - Start working on interfaces/code
  - Output structure to CDAF database