Climate processing

Organization of processing steps to generate global, daily temperature and precipitation estimates at 1km resolution over 1970-2010.

- 1 Download/preprocess data
  - LST (MOD11A1) (JR&BP)
    - aggregate-daily-lst.py ➜
      - Download tiles
      - QC flag handling, keep the following:
        - First two QA bits “00”
        - First two “01” and second two “00”
        - Explore "LST error flags" and "clear-sky coverage"
      - Calculate monthly climatologies
        - LST day for tmax
        - LST night for tmin
  - Cloud Product (MOD06_L2) (AW)
    - Download swath data (not easy due to untiled format)
    - aggregate-swath-mod06.r
      - swath->grid
      - QC flag handling
      - monthly summaries
  - Landcover
    - Jetz consensus product
      - Landcover 1 & 3 (grass & forest) (BP)
      - Add Barren, Urban, Agriculture
      - Add spatial aggregation around points (2km, 4km?)
  - Met Station Data
    - Station Data (GHCN)
      - ghcn-to-psql.R (JR) ➜
        - Download .dly files
        - Convert to tabular format
        - insert into database
      - Monthly GHCN: construct from daily or use Monthly GHCN product
    - Station Data (other)
      - If we pursue anomaly approach, we may want to use WMO climatologies
      - Literature review of (available?) data ➜
        - WorldClim data
        - South African (AW)
        - Neotropics?
        - Japan? (AW)
    - Quality Control (AW & BP)
      - Update database with QC flags
      - outliers, runs, etc
      - Adam has example script
      - Station Selection (WorldClim style priorities)
    - Spatial Query (submit polygon and date to get table of observations)
  - Other data
    - NCEP Reanalysis?
- 2 Regions ➜
  - Oregon
  - South Africa
  - East Africa
- 3 Subset
  - Spatial
    - Spatial Tiling
      - Tiling depends on methods (will we need data from surrounding tiles or just stations)?
      - Use stations from outside region to reduce edge effects
  - Temporal
    - Subset to 1970-2010
    - Day-by-day interpolations
  - Reproject
    - Use local projection (UTM?) for interpolation?
    - Fit in UTM and predict into translated cell centroids?
  - Prepare Design matrix for modelling
    - One day
    - One variable (tmax, tmin, or ppt)
    - One tile
- 4 Interpolation
  - Temporal resolution (1970-2010)
    - Daily
    - Monthly
    - Monthly Climatologies
  - Methods
    - Response variables
      - Temperature (tmax & tmin) (BP)
      - Precipitation (AW)
    - Input variables
      - Latitude
      - Longitude
      - Distance to Ocean
      - Aspect (eastness & northness)
      - Elevation
      - Land Surface Temperature (LST)
      - Cloud product?
      - Land cover
        - interactions with LST?
      - Others?
        - other terrain variables?
        - coarse gridded weather (NCEP reanalysis)
    - 'Raw' values
      - GAM
        - Input variables
          - Daily vs. monthly mean LST
        - Interactions
          - LST & Landcover
          - aspect
          - Latitude/Longitude
        - Spatial structure in error term?
      - Co-kriging
        - Variogram
          - Fit with LST data
          - Fit with station observations
          - explore various variograms (automap)
          - Bayesian?
      - GWR
        - same as above
        - range setting set (spgwr package)
        - Validation
          - Range based on training dataset
          - Krig slopes based on training data and use for predictions
    - Anomaly Approach
      - GAM
      - cokriging
      - GWR
    - Backup method (for times/locations that primary method fails)
  - Assessment
    - Literature Review: Compare to other interpolation studies (BP & AW)
      - Mendely shared library
      - Table of methods
      - Explanation of similarities and differences
      - GAM vs Spline vs ANUSPLIN
    - Validation (holdout stations)
      - metrics
        - RMSE
        - MAE
        - Bias (mean error)
        - R^2
        - Precipitation
          - sensitivity/specificity
          - Relative error rather than absolute
      - Comparisons
        - spatiotemporal distribution of metrics
        - daily vs monthly vs annual, etc.
        - median, range, quantiles
    - Model specific parameters
      - GAM
        - variance explained per variable
        - AIC
        - GCV (smoothing term)
        - RMSE (fitting dataset)
        - prediction uncertainties
      - Kriging
        - variogram parameters
        - prediction uncertainties
      - GWR
        - Summary of localized slopes
    - Compare station monthly means with other products
      - WorldClim
      - PRISM
  - Include pixel-by-pixel uncertainties?
  - parallelizing procedures
- 5 Derived Climate
- 6 Data Distribution
  - Projection ➜ Berhmann
  - Hosting posibilities
    - ➜ Use IRI's Data Library?
      - Automatic browsing, subsetting, transformation to various file formats
    - Map of life infrastructure?
  - ftp site?