TEMPO Ground Systems: Cloud Processing

Ewan O’Sullivan
John C. Houck
SDPC Software Developers

May 27, 2015
Starting point: OMI Raman-scattering cloud code (OMCLDRR)

Modification for TEMPO
  - Changes common to all science codes
  - Cloud-specific changes

Testing

Resource usage

Further work
Main products:
- Cloud Fraction
- Cloud Pressure (optical centroid pressure)

Code characteristics:
- ~10000 lines of Fortran90, ~120 subroutines / functions
- Inputs:
  - HDF-EOS2/HDF4 radiance and irradiance
  - ASCII and fortran binary format calibration files (climatologies, cross-sections, lookup tables, etc.)
- Output: HDF-EOS5/HDF5

Fast code: ~70s to process one OMI orbit.
Modifications for TEMPO: common to all codes

- **File format changes:**
  - TEMPO will use netCDF4/HDF5 throughout
  - Replace HDF-EOS with TEMPO I/O library built on netCDF4 (libtio)

- **Error handling and logging:**
  - ECS SDP Toolkit over-complicated for our needs
  - Replace with dedicated error handling and logging library (libtell)

- **Compilers:**
  - OMCLDRR built using PGI compiler
  - TEMPO software development uses GNU and Intel compilers
- Remove/disable unused experimental/test code

- Remove OMI-specific features
  - Spatial zoom mode, small pixel data, row anomaly, etc.

- Reformat fortran binary calibration files to netCDF
  - Smaller, human readable files with no significant loss of speed

- Results:
  - ~1000 lines of code removed,
  - netCDF I/O added but HDF-EOS I/O retained for testing,
  - Speed of code retained.

- Refactoring for TEMPO mostly complete
Testing

- **Regression testing**
  - Nightly build compiles and runs GNU and Intel versions of code every weekday, using OMI data
  - Output compared with baseline file “blessed” by GSFC developers
  - Cross-comparison of HDF-EOS and netCDF inputs/outputs
  - Tests specific capabilities: spline and linear interpolation, wavelength shift & squeeze, output of residuals

- **Integration testing**
  - Reduced-size TEMPO-format granule processed through Cloud, NO$_2$, HCHO codes, using pipeline manager prototype

- **Throughput tests**
  - Tests with full-size TEMPO granule - no problems identified
Resource Usage

- Time to process one OMI orbit: ~70s (Intel Xeon 2.8 GHz)
- Memory usage: ~112 MB

- Compare to other science codes:
  - Total $O_3$: ~70s
  - Trace gas (HCHO, NO$_2$): ~1 hour
  - $O_3$ profile (not yet updated for TEMPO): several hours

- TEMPO 6-minute granule size equivalent to ~2.5 OMI orbits
  - Tests confirm equivalent scale-up in processing time.

- Conclusion: Cloud code will be a minor resource user, no problems expected even with increased capabilities
Further work / TBD

- Incorporate improvements from GSFC developers:
  - Uncertainty estimation
  - Improved surface reflectance climatology

- Issues to be resolved:
  - Test data
  - Cloud mask product

- CDR-level review of Cloud, Trace gas, and Total O$_3$ codes scheduled for October
  - We expect to have a prototype pipeline including all three codes, updated for TEMPO, ready for the review

- Take-home message: Cloud code modifications for TEMPO on schedule, no problems expected