Cloud-based Linked Data Geoprocessing: Implementing Kriging as WPS on the Cloud

Elias Grinias
Department of Civil Engineering, Surveying and Geomatics, TEI of Central Macedonia

and Dimitris Kotzinos
ETIS Lab, Department of Computer Science, University of Cergy Pontoise, FR and
Institute of Computer Science, FORTH, GR
Geoprocessing on the Cloud

- Statistical Geoprocessing
  - Time and storage demanding activity

- Cloud computing
  - Efficient (distributed) solving of complex geoprocessing problems
  - Scalable and elastic management of large (input) geospatial datasets

- Linked Data (LD)
  - Data inter-linked to each other
  - Data indentified by unique URIs
  - Re-usability, interconnection and combination of data on the fly
  - Linked Open (Geo)Data: Linked (Geo)Data freely available and open on the Web
Web Processing Service (WPS)

- OGC Web Processing Service (OGC-WPS)
  - Standard for implementing/publishing geo-processes as a web service
  - Operations
    - GetCapabilities, DescribeProcess, Execute
    - In practice: HTTP XML submission to service’s URL
  - Execute operation
    - XML containing information about
      - Process’ identifier
      - Inputs/parameters and form of process’ outputs
Implementation Schema

- Geoprocessing on the cloud
  - WPS based on 52 North WPS implementation
  - Kriging WPS process based on R-Gstat
  - Input fetched from Linked Open GeoData repository through (Geo)SPARQL queries posted to Virtuoso server.
  - Usage of more processing and storage instances when needed.

- Client Implementation
  - Modification of 52 North WPS Client Implementation
  - Based on latest javascript open source software: Ext, Openlayers and JQuery modules.
Kriging on the Cloud
Ordinary Kriging

- Geostatistical interpolation method
- Fact: as distance between points increases, the similarity, defined by the covariance or correlation between points, decreases.
- Local neighborhood prediction: predict the unknown value at a location $x_0$ using data values in the neighborhood of this location.
Kriging on the Cloud Implementation

- Adoption of 52 North WPS 3.1.1 implementation of OGC-WPS 1.0.0 standard
- Ordinary kriging using R-Gstat
  - Global or local-neighborhood prediction
  - Prediction on non-projected data using great circle distance between points
  - Fast enough (its main functionality is coded in C)
- Interconnection between R and the Java module located at the WPS Container through TCP/IP server Rserve.
Client Implementation

WPS Process Description

- Modification of the open source 52 North Openlayers WPS Client

WPS Process Description

- Openlayers panel construction after an HTTP GET request to WPS Container for Kriging inputs/parameters and output description
Client Implementation
Vector Data Loading on Map

- Openlayers Vector Layers
- Data in CRS WGS84 (non projected data)
- Layer sources:
  - User’s own data in Excel format
  - Selection box on already loaded layers
  - Linked GeoData fetched from Virtuoso server using fixed SPARQL queries
  - Linked GeoData fetched from Virtuoso server using partially parameterized SPARQL queries
Client Implementation
Kriging Execution

- Use of the panel for selecting
  - Input layer
  - Kriging parameters
to construct the XML for the Execute operation
- Execute operation: XML transfer to WPS Container through an HTTP POST request

Preview in PNG Format (one of process’s output)
Cloud implementation

- Using standard AWS load balancing
- New instance generated when CPU load reaches >75%
- Use of a shared/distributed file system (GlusterFS)
Linked Open GeoData

● Linked Data (LD)
  ● Data inter-linked to each other
  ● Data identified by unique URIs
  ● Re-usability, interconnection and combination of data on the fly

● Linked Open (Geo)Data
  ● Linked (Geo)Data freely available and open on the Web
Linked (Open) Data as a Service

- Abstraction layer for data access
  
  *abstract the applications from the specific setup of the data management service (such as local vs. remote, federation, and distribution)*

- Beyond Data Access
  
  - Enabling automation of discovery, composition, and use of datasets
  
  - Data Markets
  
  - Online Visualization Services
  
  - Data Publishing Solutions
  
  - Data Aggregators
  
  - BI / Analytics as a Service
Linked Data Services

- **addR2RMLMappings**: Structured data (XML documents)
- **addXSLMappings**: Structured data (relational tables)
- **ldquery**: Geolquery
- **ldexport**: GeoJSON, KML, ShapeFile, ...
- **formats**: xml, tsv/csv, json, trig, trix, n3, ...
- **ldimport**: Inspire compliant
- **inspire_export/inspire_query_export**: GeoJSON Infrastructure for Europe
Cloud implementation

- Using standard AWS load balancing
- New instance generated when CPU load exceeds 70%
- Use of Virtuoso RDF triplestore
- REST API calls are posted independently by each WPS instance
Conclusions

- Geoprocessing on the cloud
  - WPS based on 52 North WPS implementation
  - Kriging process based on R-Gstat
  - Input fetched from Linked Open GeoData repository
  - Usage of more processing and storage instances when needed
Future Work

- Addition of more geo-processes in WPS
- Full parallelization of geoprocessing algorithms, when possible
- Adoption of kriging algorithms under the MapReduce computing paradigm

- Work presented here is part of the InGeoClouds Project
  www.ingeoclouds.eu
Questions?

Dimitrios.Kotzinos@u-cergy.fr