FOCUSED STUDY ACTIVITY WORK PLAN

General Information

Integrated geospatial environmental analytics of OSM monitoring data.		
Work Plan Unique Identifier:	E--6-1718	
Focused Study Activity Title:	Big Data Environmental Analytics – Integrated Geospatial Environmental Analytics of OSM Monitoring Data.	
Focused Study Category:	Investigation of Cause or Potential Ecological Impact	
Geographic Location (*choose from drop-down menu. If Project Location is in more than one area choose from second drop-down***)**	Athabasca Oil Sands Region	Athabasca Oil Sands Region
Monitoring Site(s) Coordinates *(latitude and longitude)*	This project does not involve any environmental monitoring in 2017-18. On contrary, this project analyzes the monitoring data collected in past under JOSM/OSM.	
Project Leader:	Anil Gupta	
Organization and contact information:	Environmental Monitoring and Science Division Alberta Environment and Parks Email: anil.gupta@gov.ab.ca Tel: 403 297 3930	
Date Study initiated:	April 1st 2017	
Monitoring Category: *(From OSM long-term plan; choose from drop-down menu)*	Watershed Monitoring	
Strategic Objective of Focused Study: *(From OSM long-term plan; choose from drop-down menu)*	Objective W3: Integration and Synthesis	
Hypotheses: *(Briefly outline the specific hypotheses that your focused study is aiming to address)*	Investigation of Cause or Potential Ecological Impact: As this project will deal with all of the JOSM aquatic data, several individual or combined (generalized) hypotheses could be tested. Following integrated composite hypothesis will be tested under this	

	project: Ho1: Measured environmental endpoints are not correlated with each other and also with other measurements of the ecosystem including geo-physical characteristics of the basin.
Deliverables: *What tangible goal (s) and/or product(s) will the monitoring produce and when?*	The project will produce following tangible outputs: 1. Compilation of pre and post OSM aquatic monitoring data sets into a standardized machine readable digital format (The compile data sets will be linked to partner federated data systems or OSM data portal when developed). 2. Geospatial database of key physical data sets e.g. historical land use, land use changes, climate change, contributing watershed area at key monitoring locations, soil distribution in contributing watershed area, watershed characteristics, drainage features, surficial and bedrock geology, soil types etc. 3. A water quality data analysis tool (most likely in R) 4. Integrated data analysis to identify trends, patterns and "hot spots" and also to inform cumulative effects assessments.

Detailed Study Plan

(Please provide detailed information on the specifics of your focused study including – (**keywords, hypothesis and the assumptions and constraints behind your hypothesis**)

Provide a maximum of 10 key words that describe this project. Use commas to separate them:

Environmental analytics, cause and effects, environmental impacts, aquatic system, Environmental predictions, climate change, land use change, Athabasca river, oil sands, neural networks, artificial intelligence, fuzzy logics, regression.

Describe how you will test your hypothesis:

Project Description: Most regulatory bodies in North America have established an extensive and comprehensive program for monitoring the environment. Under the Canada-Alberta Joint Oil Sands Monitoring Plan (JOSMP), Alberta also has large networks of monitoring stations to collect various environmental data including physical, ecological and biological (e.g. hydrology, meteorology, water quality, benthic invertebrates, fish health and communities, wildlife health, atmospheric depositions, air quality, terrestrial etc.) These data can be used to understand changes in the environment due to natural/anthropogenic activities. While we are good at identifying variations in the short term (that too in a particular sub-component), slow creep in data over longer terms are generally missed. If long term data over a wide geography are available there is an opportunity to do spatial & temporal trend analysis and establish possible causal relationships, for example, between changes in water quality parameters and those that may impact it such as land use changes, meteorological factors and human activities in the vicinity. Using different modeling techniques statistical as well as other newer techniques (e.g. neural networks, artificial intelligence, fuzzy logics etc.), the environmental data can be analyzed.

The Scientific Integrity Expert Panel Review also highlighted the limited integration of activities and reporting within and across the four monitoring components and emphasized the need of comprehensive data analysis across all four monitoring components (air, water, land-biodiversity & wildlife) to better understand and support the evaluation of the ecosystem changes. Assessment of the pre-JOSM data is needed to provide a historical basis for comparison and extend the data time series to better understand long-term trends and changes in the ecosystem.

Background and Rationale*:* Under the Canada-Alberta Joint Oil Sands Monitoring Plan (JOSMP), large amount of various environmental data has been collected since inception of JOSMP in 2012. Considerable efforts were made to interpret and synthesize water data in last fiscal years (2015-16 and 2016-17). However, all the collected have yet to be synthesized and interpreted in an integrated manner (within water component and across other component) to identify patterns & trends and establish possible relationships between trends, environmental events and changing ground conditions. The Scientific Integrity Expert Panel Review (2016) also highlighted the limited integration of activities and reporting within and across the four monitoring components and emphasized the need of comprehensive data analysis across all four monitoring components (air, water, land-biodiversity & wildlife) to better understand and support the evaluation of the ecosystem changes. Further, the review panel also recommended the assessment of the pre-JOSM data to provide a historical basis for comparison and extend the data time series to better understand long-term trends and changes in the ecosystem.

To address the panel's recommendation, the proposed work plan involves an integrated data syntheses of key environmental data sets collected under JOSMP and where possible extending the data series by combining the pre-JOSMP environmental monitoring data.

Literature Review**:** In order to develop appropriate methodology for hypothesis testing, a comprehensive literature review has been conducted.

Hirsch et al. (1982) used three statistical techniques (Seasonal Kendall test for trend, Seasonal Kendal slop estimator, and flow adjustment coupled with the seasonal Kendall test) to analyze the monthly water quality data. These techniques identified significant monotonic changes in the water quality parameters as well as relationships between parameter values and water discharge. Pattern recognition techniques based on multivariate statistics such as factor analysis (FA), cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA) and neural networks have widely been used to analyze large water-quality data matrices (Antonopoulos et al., 2001; Singh et al., 2004; Vega et al., 1998; Yidana et al., 2008; Ha and Stenstrom, 2003; Singh et al., 2009; Zhang and Stanley, 1997; Zhang et al., 2002). Alberto (2000) used multivariate statistical methods such as (CA, FA/PCA and DA) to evaluate the spatial and temporal variations of water quality in Suquia River basin (Cordoba-Argentina). The results showed that CA can identify spatial and temporal differences but failed to provide details of the difference. FA/PCA grouped the selected parameters according to common features. DA gave the best results for both temporal and spatial analysis.

In a different study, Shrestha and Kazama (2006) assessed the surface water quality of Fuji river basin using multivariate statistical techniques. The analysis of eight years water quality data indicated that CA can help develop an optimal sampling strategy such as reducing the number of sampling stations and the parameters; FA/PCA can help extract and identify the factors/sources responsible for variations in river water quality; their results also found that DA was the best technique for both spatially and temporally analysis of water quality data. Yidana (2008) applied CA and PCA simultaneously to surface water hydro-chemical data from three different locations and established the source of variance in the water quality data in three different locations in Ankobra Basin.

A neural network is a computational tool which mimics the behavior of the human brain. It is a black-box model, which does not require information of the internal functions of a system in order to recognize relationships between inputs and outputs (Ha and Stenstrom, 2003; El-Din and Smith, 2002; Schleiter et al, 1999). Singh et al (2009) found that neural networks can be an effective tool for the computation of river water quality and can be used to understand the trend in river pollution. Zhang and Stanley (1997) used it to predict the water quality in North Saskatchewan River. Ha and Stenstrom (2003) successfully developed a neural network models to identify different types of land use based on six years of storm water quality data.

<u>Hypothesis testing methodology:</u>

The research goal of this proposal is to analyze long term environmental data as a whole to identify patterns & trends and establish possible relationships between trends, environmental events and changing ground conditions. It is understood that there is some research done to identify the trends, but these are done only with localized information on an isolated data sets and never considering multiple data sets across four monitoring components and also from multiple monitoring locations (spatial scale). To achieve this goal, the following specific objectives are identified:

- Compilation of pre- and post JOSM environmental monitoring data sets - this will include water quality, meteorology and hydrology. Where possible, this project will also attempt to incorporate other ecological and biological data sets to establish intrinsic inter-relationships among various sub-components of environment.
- Compilation of physical data sets e.g. historical land use, land use changes, climate change, watershed area, drainage features, surficial and bedrock geology, Soil types etc.
- Assessment of compiled data - this will include understanding the quality/reliability of data, data history, identify missing (or erroneous) information, the geography of information distribution and time lines over which the data has been collected, data completeness and suitability for inclusion in assessment.
- Analysis of environmental data to identify patterns and “hot spots” where certain parameter values are relatively high. Analyze the data further to establish “cause” for hot spots and possible consequences. Statistical methods will be used to analyze for hot spots and possible consequences. Hot spot trends with time and changing ground conditions will be identified (whether natural or anthropogenic) and tested. “What if” scenarios will be hypothesized and hypothesis testing statistics used to test for them.
- Analyze patterns of variation of key water quality parameter with season (time), location (geography), land use activity, proximity to water bodies and meteorology. Established pattern recognition techniques will be used to identify data patterns and establish possible relationships.

Principle component analysis will be used to establish orthogonality of information and understand co-dependence of information.
- Make recommendations for adaptive monitoring program in future.

Assumptions and Constraints behind the hypothesis and the testing method:

The extensive literature review conducted prior to conceiving this project strongly suggests validity of hypothesis and tremendous value in testing the hypothesis to draw insights about relationships between numerous monitoring variables and investigation of cause-effects that could lead in formulating predictive relationships. This project will also apply some newer analysis techniques with assumption that these techniques are also well suited for environmental data. Therefore, we have reasonably good confidence in the success of the project. However, environment is a highly complex system of interconnected sub-systems which are not separable. Therefore, in order to get reasonable insights of this complex system, an integrated environmental analysis (across all interconnected sub-systems of the environment) is required. However, the current focus is on aquatic system and in absence of other sub-systems, it might poses some challenges in investigation and interpretation of cause-effect relationships that exists and influenced by other sub-systems not being included in present study. The whole purpose of this project is to start filling the science knowledge gap.

References:

Alberto, W.D., del Pilar, D.M., Valeria, A.M., Fabiana, P.S., Cecilia, H.A. and de los Ángeles, B.M., 2001. Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study: Suquía River Basin (Córdoba–Argentina). *Water Research*, *35*(12), pp.2881-2894.

Antonopoulos, V.Z., Papamichail, D.M. and Mitsiou, K.A., 2001. Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece. *Hydrology and Earth System Sciences Discussions*, *5*(4), pp.679-692.

El-Din, A.G. and Smith, D.W., 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. *Water Research*, *36*(5), pp.1115-1126.

Ha, H. and Stenstrom, M.K., 2003. Identification of land use with water quality data in stormwater using a neural network. *Water Research*, *37*(17), pp.4222-4230.

Hirsch, R.M., Slack, J.R. and Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research*, *18*(1), pp.107-121.

Schleiter, I.M., Borchardt, D., Wagner, R., Dapper, T., Schmidt, K.D., Schmidt, H.H. and Werner, H., 1999. Modelling water quality, bioindication and population dynamics in ecosystems using neural networks. *Ecological Modelling*, *120*(2), pp.271-286.

Shrestha, S. and Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, *22*(4), pp.464-475.

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. and Kouimtzis, T.,

2003. Assessment of the surface water quality in Northern Greece. *Water research*, *37*(17), pp.4119-4124.

Singh, K.P., Basant, A., Malik, A. and Jain, G., 2009. Artificial neural network modeling of the river water quality—a case study. *Ecological Modelling*, *220*(6), pp.888-895.

Singh, K.P., Malik, A., Mohan, D. and Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Research*, *38*(18), pp.3980-3992.

Vega, M., Pardo, R., Barrado, E. and Debán, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, *32*(12), pp.3581-3592.

Reporting and Publications

Provide information on the anticipated reports / publications. (Insert additional rows if needed)

Expected Subject/Titles of Publications or Reports	Short Description of Publication or Report	Expected Year of Publication
Environmental Big Data Analytics -Technical Report Series	This report will be one of the reports in the intended Technical Report Series under OSM. The report will describe the availability and suitability of collected OSM environmental data, formulated hypothesis and testing methodologies applied with discussion on the analytics results including trends, hot spots, spatial and temporal variations, predictive and other relationships (if exist). The report will also identify the data gaps and provide recommendations for future monitoring in the oil sands area.	2018-19
Journal publication	Depending upon the outcome of analysis and result interpretation, a peer reviewed journal publication is being aimed to disseminate the findings.	2018-19

Technical / Professional Roles and Responsibilities

Identify members of the monitoring team/organization, their roles and responsibilities. Identify monitoring organization leads if different from overall monitoring activity lead. (Insert additional rows if needed)

Role	Responsibilities	Resource Name/Organization
Project Lead and PI	Overall project management; Supervising the graduate student/post doc or RA at university, Develop and coordinate deliverables in the work plan.	Anil Gupta (AEP)
Co-PI	Co-supervising the graduate student/post doc or RA at university	Gopal Achari, PhD, PEng., FCSCE Associate Dean, Graduate Studies, Schulich School of Engineering Director, Centre for Environmental Engineering Research and Education (CEERE) Professor, Department of Civil Engineering Schulich School of Engineering University of Calgary
Scientific data analysis Support	Undertake preparation, data analysis and reporting	Research Associate (already hired) at the University of Calgary

Deliverables (Year 1) If your Focus Study is longer than 1 year then complete **Appendix C** for multi-year deliverables breakdown

Provide a summary of tangible quarterly deliverables. Identify major project areas (deliverables) and results that can be identified as a tangible goal. This could include: field work, lab work/ analysis, evaluation, data, reports, publications, SOPs etc. Do not define process as your Deliverable e.g. 'fly to Ft. McMurray to conduct fieldwork' or 'seek Director approval for report'.

Deliverable(s) (please provide enough information to support status reporting)
Q1 – April to June
Literature search and finalization of hypothesis testing methodology
Compilation of pre and post OSM aquatic monitoring data sets into a standardized machine readable digital format (The compile data sets will be linked to partner federated data systems or OSM data portal when developed).
Q2 – July to September
Assessment of compiled data - this will include understanding the quality/reliability of data, data history; identify missing (or erroneous) information, data completeness and suitability for inclusion in assessment.
Compilation of key physical data sets e.g. historical land use, land use changes, climate change, contributing watershed area at key monitoring locations, soil distribution in contributing watershed area, watershed characteristics, drainage features, surficial and bedrock geology, soil types etc.
Q3 – October to December
Developing a data analysis tool (most likely in R)
Integrated data analysis to identify trends, patterns and "hot spots" and also to inform cumulative effects assessments.
Q4 – January to March
Investigation of cause-effect relationships and result interpretation.
Paper - manuscript

Detailed Financial Breakdown – Year 1 of 3 (2017-2020)

Also complete **Appendix B** for the multi-year financial breakdown

Budget requirements – List areas that require budget expenditures: (ADD OR DELETE BUDGET CATEGORIES AS REQUIRED)	**OS Funding**	**External Funding (*outside JOSM*)**
O&M - Operations and Maintenance:		
Helicopter Costs	**$**	**$**
Field Costs	**$**	**$**
Data Management	**$**	**$**
Internal Lab Analysis	**$**	**$**
Consumable Materials & Supplies Including software/hardware	**$10,000**	**$**
Sub-Total	**$10,000**	**$**
O&M - Travel		
Field Work	**$**	**$**
Conferences (*identify conference*) Attendance to conferences to present research findings Publications	**$7,000** **$2,500**	**$**
Meeting (*identify meeting*)	**$1,000**	**$**
Sub-Total	**$10,500**	**$**
O&M - External Contracts :		
Goods and Services Contract (*describe contractor*)	**$**	**$**
External Lab Analysis	**$**	**$**
Sub-Total	**$**	**$**
Salaries:		
Principal Investigator – Anil Gupta (0.25 PY)		**$**
Technical / Professional Assistants 1. Post Doc or Research Associate	**$75,000**	**$**
Field Staff	**$**	**$**
Sub-Total	**$75,000**	**$**

Budget requirements – List areas that require budget expenditures: (ADD OR DELETE BUDGET CATEGORIES AS REQUIRED)	**OS Funding**	**External Funding** (*outside JOSM*)
Total Salaries	**$75,000**	**$**
Total O&M	**$20,500**	**$**
2017-2018 GRAND TOTAL*	**$95,500**	**$**

Appendix A - Approvals

Project Submitted by:

Name: Anil Gupta		
Organization: Alberta Environment & Parks	Signature:	Date: 11 Aug 2017

Project Approved by:

Dr. Monique Dubé (AEP)		**Dr. Kevin Cash (ECCC)**
Signature		Signature
Date		Date

APPENDIX B – Detailed Multi-year Financial Breakdown (Complete the following detailed financial breakdown; add or delete categories as required)

Budget requirements	Year 1 (2017- 2018)		Year 2 (2018- 2019)		Year 3 (2019- 2020)	
	Cash	In-kind	Cash	In-kind	Cash	In-kind
1) Salaries and benefits						
a) Investigators						
b) Technical/professional assistants	$75,000		$75,000		$75,000	
c) Field Staff						
d)						
2) Operations and maintenance						
a) Facilities						
b) Equipment						
c) Lab analysis						
d) Data management						
e) Field work						
3) Consumable Materials and supplies						
a) Software/hardware	$10,000		$10,000		$10,000	
b)						
4) Travel						
a) Conferences and meetings	$7,000		$7,000		$7,000	
b) Field work						
c) Project-related travel	$1,000		$1,000		$1,000	

5) Dissemination & Engagement						
a) Publications/Reports	$2,500		$2,500		$2,500	
b) Translation (if required)						
c) Communications						
d) Stakeholder Engagement						
e) Indigenous Peoples Engagement						
6) External Contracts						
a)						
Grand Total	$95,500		$95,500		$95,500	

APPENDIX C –Years 2 and 3 Deliverables (Complete the following detailed breakdown. Provide a summary of tangible quarterly deliverables and your anticipated expenditures. Identify major project areas (deliverables) and results that can be identified as a tangible goal.)

Year 2 (2018- 2019)
Deliverable(s) (please provide enough information to support status reporting)
Q1 – April to June
Report on 2017-18 work (published as JOSM Technical Series)
Paper - manuscript
Q2 – July to September
Literature search to identify additional innovative techniques for hypothesis testing and accordingly adjusting the methodology
Update to aquatic monitoring database to include 2017-18 monitoring data
Compilation of other relevant monitoring data (e.g. dry and wet depositions, wetlands, landscape disturbances etc).
Q3 – October to December
Advance geospatial analysis and update to geospatial database of key physical data sets e.g. historical land use, land use changes, climate change, contributing watershed area at key monitoring locations, Land use patterns and soil distribution in contributing watershed area, watershed characteristics, drainage features, surficial and bedrock geology, soil types etc.
Q4 – January to March
Advanced geospatial analysis of environmental data to identify trends, patterns and “hot spots” and analyses of variation patterns – by applying other innovative techniques.
Paper - manuscript

Year 3 (2019- 2020)
Deliverable(s) (please provide enough information to support status reporting)
Q1 – April to June
Report on 2018-19 work (published as JOSM Technical Series)
Paper - manuscript
Q2 – July to September
Literature search to identify additional innovative techniques for hypothesis testing and accordingly adjusting the methodology
Update to aquatic monitoring database to include 2018-19 monitoring data
Compilation of other relevant monitoring data (e.g. air and biodiversity etc.).
Q3 – October to December
Advance geospatial analysis and update to geospatial database of key physical data sets e.g. historical land use, land use changes, climate change, contributing watershed area at key monitoring locations, Land use patterns and soil distribution in contributing watershed area, watershed characteristics, drainage features, surficial and bedrock geology, soil types etc.
Q4 – January to March
Advanced geospatial analysis of environmental data to identify trends, patterns and “hot spots” and analyses of variation patterns – by applying other innovative techniques.

Paper - manuscript
Report on 2019-20 work (published as JOSM Technical Series)