Université du Québec
Institut national de la recherche scientifique

Agriculture et Agroalimentaire Canada | Agriculture and Agri-Food Canada

Evaluating the propensity of saturation excess runoff using a topographic index (wetness index) with NHN and DEM GeoBase data

Authors:
Jean-Thomas Denault
Alain N. Rousseau
Eric van Bochove
Farida Dechmi

December 2008

Canada

Agriculture and Agri-food Canada/Agriculture et Agroalimentaire Canada
2560 boul. Hochelaga, Quebec (Quebec), G1V 2J3
CANADA

Institut National de Recherche Scientifique-Centre Eau, Terre et Environnement
490 de la couronne.Quebec (Quebec) G1K 9A9
CANADA

Table of Contents

Use Case Overview

Saturation-excess runoff is one of the key processes involved in hydrology of temperate and humid climate. It is usually driven by topography for areas with gentle slope and shallow soils on impermeable bedrock or an impervious soil layer. In the context of pesticides, pathogens or nutrients-rich soils, knowledge of soils prone to saturation-excess runoff can be used to illustrate the spatial distribution of watershed areas at risk of overland agricultural pollutant transport to surface water.

Agriculture and Agri-Food Canada (AAFC) has initiated the development of agri-environmental indicators to assess agriculture performance. Indicators of Risk of Water Contamination by Phosphorus and by Coliforms (IROWC-P and IROWC-Col.) were developed using an index approach integrating sources, transport and connectivity factors. Saturation excess runoff was accounted for as one element of connectivity linking agricultural land to the water bodies. To evaluate this process, a Topographic Index (TI) based on the TOPMODEL approach was computed for all agricultural watersheds of Canada using GeoBase CDED 1:50 000 digital elevation models (DEMs) and their corresponding National Hydro Network (NHN) GeoBase data layers. All algorithms were integrated in GIS software developed by Institut National de Recherche Scientifique-Centre Eau, Terre et Environnement: PHYSITEL. Mean TI values of agricultural land were used to classify agricultural regions.

Software and Data Sources

To evaluate the risk of saturation excess runoff throughout Canada, Topographic Index (TI) matrices were calculated using digital elevation models (DEMs) and Hydro Network geospatial vector data. GeoBase CDED at 1:50 000 scale available at Geobase (www.geobase.ca) and their corresponding National Hydro Network (NHN) GeoBase layers were the source data actually used. All pre and post data processing were implemented using ArcGIS-ARC/INFO 9.2 (www.ESRI.com). All network data format transformations were done using FME 2006 (www.safe.com) and all other calculations were performed by PHYSITEL (v.3.0) [Turcotte *et al*., 2001; Royer *et al*., 2006; Rousseau *et al*., 2009]. References in this use case are based on the software used.

The Challenge

Agriculture and Agri-Food Canada (AAFC) has initiated the development of agri-environmental indicators to assess agricultural Beneficial Management Practices (BMPs) efficiency and to provide objective, science-based assessment of agriculture sustainability[1]. The Indicator of Water Contamination by Phosphorus (IROWC-P) and the Indicator of Water Contamination by Coliforms (IROWC-Col.) were both developed using an index approach integrating characterization of source and transport processes.

Indicator values = Contaminant Source x Transport_Hydrology (1)

Transport components account for surface runoff, drainage and soil erosion while Hydrology components involve all important connectivity factors that facilitate the transport of contaminant between agricultural land and waterbodies. Connectivity factors include surface drainage density, artificial drainage, likelihood of developing preferential flow and propensity of developing saturation excess runoff.

Saturation excess runoff is one of the key hydrological processes in temperate and humid climate. This kind of runoff is usually driven by topography for areas with gentle slope and shallow soils on impermeable bedrock or an impervious soil layer (Figure 1). The surface runoff process can be predicted with an index of hydrological similarity based on topographic considerations, the Topographic Index (TI) introduced by Beven and Kirkby (1979). Following this concept, all topographic units or spatial elements of a watershed with an identical index value develop, in principle, the same conditions for saturation, surface and subsurface flow/runoff.

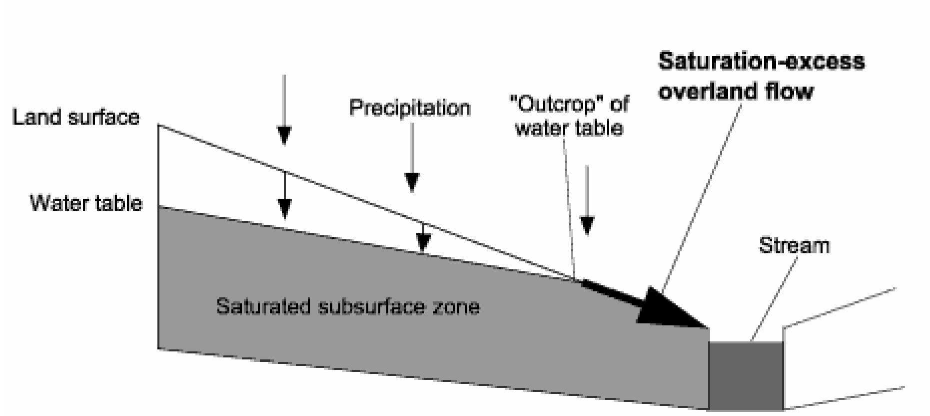


Figure 1. Saturation excess overland flow.

The TI connectivity matrices for Canada were developed in collaboration between the Institut National de Recherche Scientifique (INRS) - Centre Eau, Terre et Environnement,Agriculture and Agri-Food Canada (AAFC), and DelDegan, Massé Experts-conseils.

[1] National Agri-Environmental Health Analysis and Reporting Program, http://www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1181580464260&lang=e

The Solution

Conceptual Considerations

TI values are evaluated for each pixel of a grid based on equation 2.

$$TI_i = \ln\left(\frac{a_i}{\tan(\beta_i)}\right) \qquad (2)$$

TI_i is the topographic index for pixel i. High values will be caused by either long slope or upslope contour convergence and low slope angles, and the corresponding areas will tend to saturate first. *a* is the upslope area per unit contour length draining through the pixel i. This area is calculated using a flow accumulation algorithm along the watershed terrain under saturated conditions. By comparing a set of flow direction algorithms published in the literature, it appeared in our preliminary work that the *D8-Least Transversal Deviation* [Orlandini *et al*., 2003] is the most suited algorithm for the determination of the upslope area [Hentati *et al*., 2005]. However, this algorithm doesn't define clearly, like many others, flow directions in flat areas [Garbrecht and Martz, 1997] which often make the computation of TI difficult. Consequently, this eliminates the exact match between modelled flow directions and actual river network location. To bypass this problem, that is to correct the modelled flow directions, Turcotte *et al*. [2001] developed a new method using a digital river and lake network (DRLN) as input in addition to the DEM. The use of an initial reconditioning of a DEM by the "burning" procedure to force drainage in the most accurate position of streams in the landscape is necessary, especially in flat areas or around lakes. The hydraulic gradient of the saturated zone can be approximated by the local surface topographic slope measured with respect to plan angle (in degree) tan(β).

Since TI values are highly influenced by DEM resolution and by flow-routing algorithm, the use of uniform source information was necessary. Geobase data provided the uniformity needed for river network data through NHN and DEM data, both derived from federal 1:50 000 topographic data (re: National Topographic Data Base) or provincial/territorial data.

To be used in the evaluation of the IROWC-P and IROWC-Col, local TI values needed to be upscaled to a regional value. To do so, an agricultural vector mask representing area of agricultural production produced by Agriculture and Agri-Food Canada was used to aggregate local TI values.

Methods

The extent of the agri-environmental indicators being of national scale, TI values were calculated for all watersheds overlapping an agricultural region. Limitations in the size of the watersheds based on computing memory required the identification of all single outlet watersheds of a maximum size of approximately 10 000 km^2. A selection of agricultural watersheds was done in order to import all available data from GeoBase CDED DEM 1:50 000 and NHN indexes.

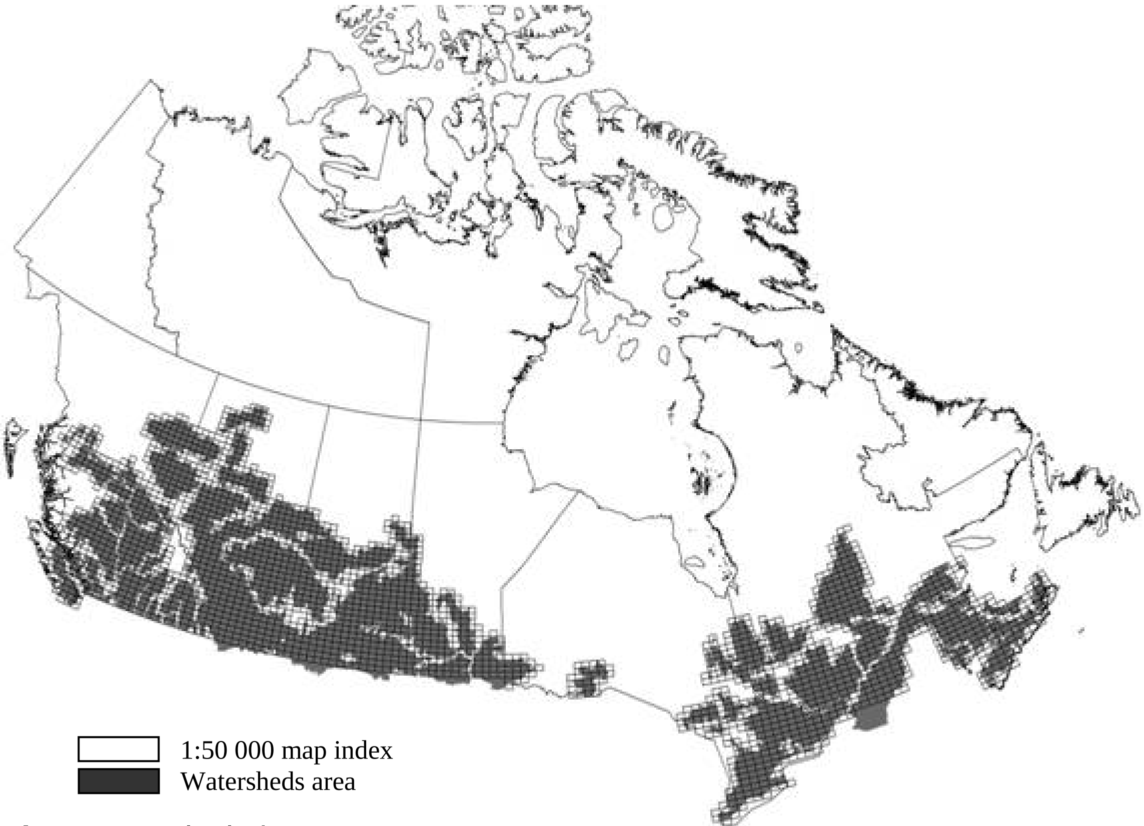


Figure 2. Calculation extent map

1. Pre-Treatment of Input Data

To have a uniform set of parameters and unique geospatial base, all input data were referenced using the Canadian Lambert Conic Projection System described below.

Projected Coordinate System: Canada_Lambert_Conformal_Conic
Projection: Lambert_Conformal_Conic
False_Easting: 0.0
False_Northing: 0.0
Central_Meridian: -96.0
Standard_Parallel_1: 50.0
Standard_Parallel_2: 70.0
Latitude_Of_Origin: 40.0
Linear Unit: Meter

A. DEM Preparation

An index of 1:50 000 mapsheets (Figure 2) was generated to select the DEM files required for the creation of each watershed DEM. This index served in the automation of the mosaic generation process made using ARC/INFO algorithms. Resulting DEMs were then projected in the Lambert Conic Projection (output coordinate system) and resampled using bilinear interpolation (resampling technique) to create models with a constant pixel size of 21 m (output cell size). This preprocessing and its underlying operations are summarized in the following diagram (Figure 3). Example of the necessary steps needed when assembling the DEM can be found in different articles published on ESRI website [2,3].

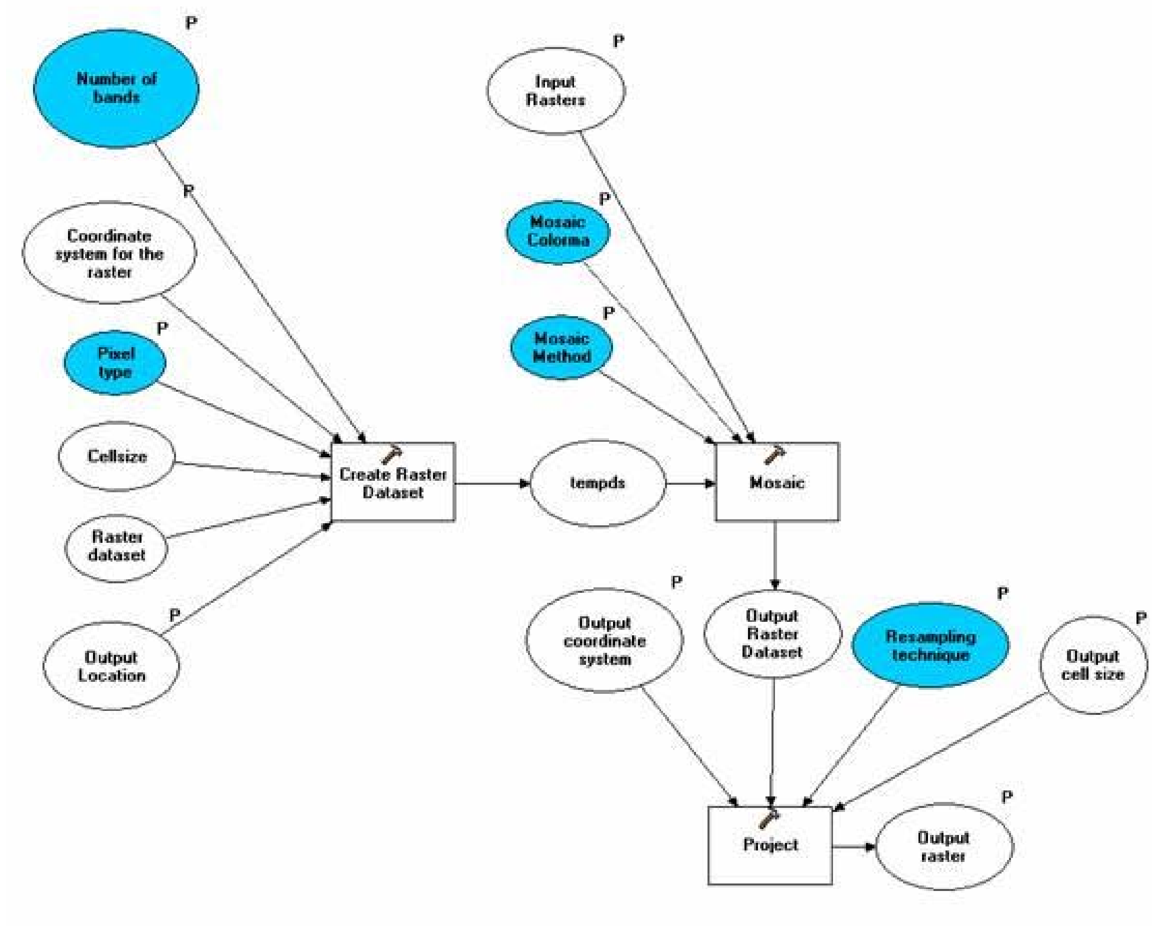


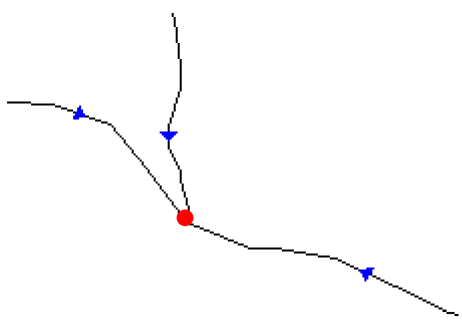
Figure 3. DEM mosaic and reprojection process

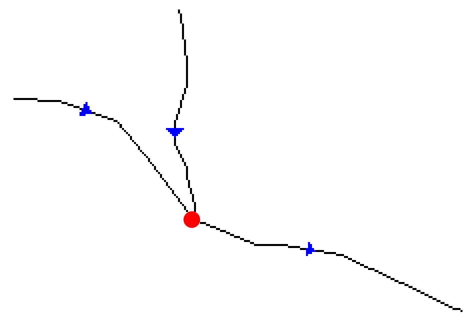
B. DRLN Preparation

Organizing the digital river and lake network (DRLN) data implies the verification of major topological constraints as listed below.

i. All river segments should be represented by single line features digitized according to flow direction.

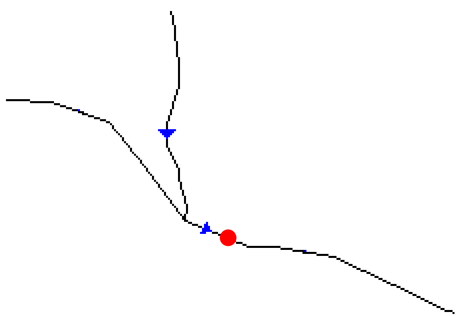
[2] http://gis.esri.com/library/userconf/proc00/professional/papers/PAP731/p731.htm
[3] http://www.esri.com/news/arcuser/0701/moredem.html

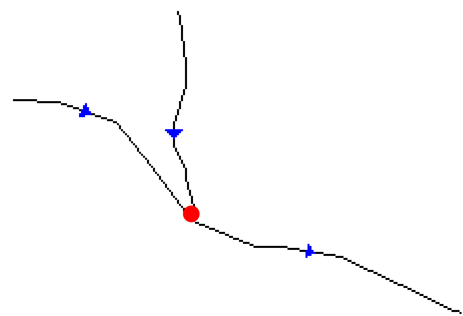




Incorrect   Correct

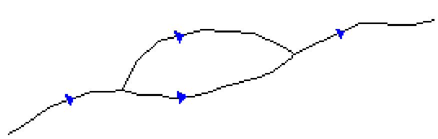
ii. Convergent streams should be connected on a single node

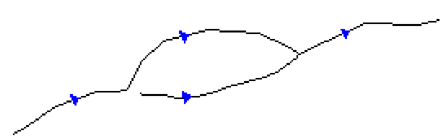




Incorrect   Correct

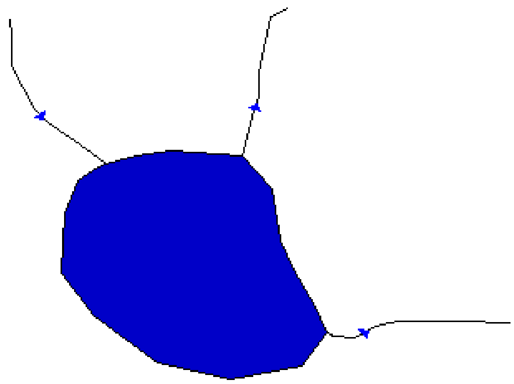
iii. DRLN does not allow divergent river flow. Major route need to be selected.



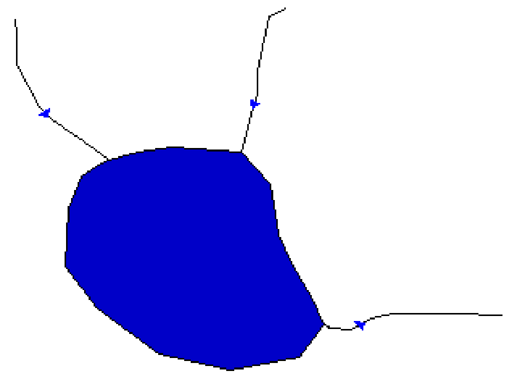


Incorrect   Correct

iv. All waterbodies should have a single outlet. (The latest version of PHYSITEL allows multiple outlet polygons. However, to simplify the calculation, single outlet polygons are preferred.)



Incorrect



Correct

Using NHN Network Linear Flow (NLF) feature data provided a high quality input data respecting all previously mentioned topology constraints. Because of the large extent of the work involved, NHN data availability (national coverage) combined to the high quality of the product saved invaluable time that otherwise would have been needed for topology corrections.

i. Network Linear Flow (NLF) Selection

The first step in preparing DRLN data was to create a linear network digitized according to flow direction. In order to select the main or a unique route flowing downstream, a geometric network needs to be generated first using the NHN *Network Linear Flow* (NLF) feature file provided for each drainage area. A simple SQL query in ArcGIS allows for the selection of all NLF features that have an "Unknown" (-1) or "Secondary" (2) value for the Level Priority attribute, which is meant to classify the NLF route within the hydrographic network.

- Using select by attributes:
  Select FROM NHN_XXXX_HN_NLFLOW_1.shp WHERE PRIORITY =’2’ or PRIORITY =’ -1’

- Open the attribute table of the NLF features and using field calculator tool
  Set all selected values to ENABLE = ‘False’

In ArcGIS, it was then easy to "disable" the selected features which resulted in allowing the determination of arc flow direction. After placing a flag at the outlet, a selection of all the lines obtained from the *Trace Upstream* tool from the ArcGIS Network Utility (or Utility Network Analyst) toolbar was performed. To do so, the analysis results should be represented as a selection of features (Figure 4).This selection representing all rivers flowing to the outlet was then saved in a separate shapefile by exporting the selected features (*WSD_ID_R.shp*).

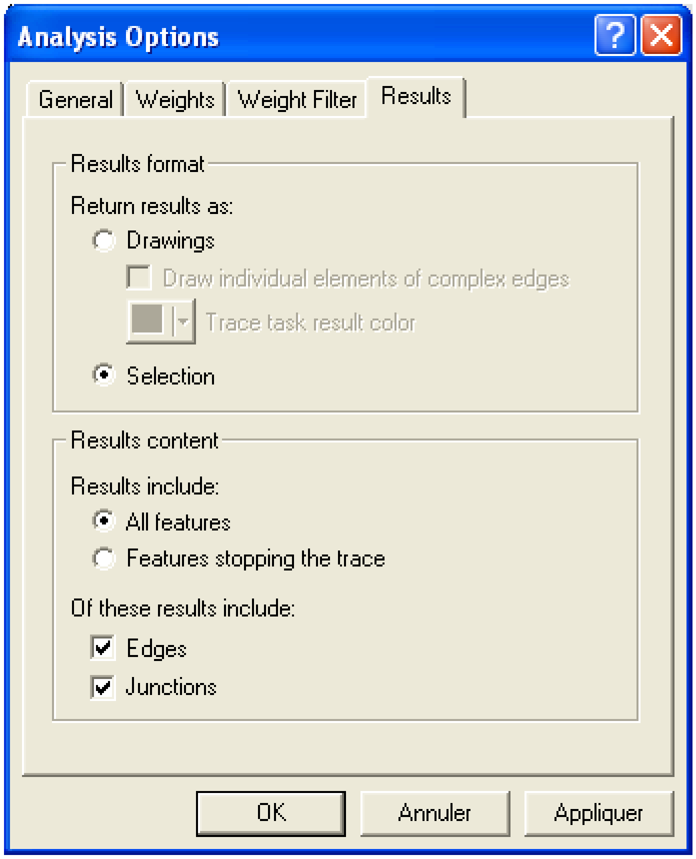


Figure 4. Network utility toolbar analysis options

ii. Organizing the River-Lake Network

All the lakes in the DRLN are represented by polygons. Note that only lakes directly connected to the main river flow were considered. The next step was to select all polygon features with known flow directions. All isolated features were disregarded for the DRLN. Using Waterbody feature polygons included in NHN data (Re: NHN Feature Catalogue) is highly recommended since this data respects all topology requirements mentioned earlier under section B.

To select the connected waterbodies, perform a selection by location where waterbodies intersect the main flow arcs saved in the previous step (WSD_ID_R.shp). The selection was performed in ArcGIS and then saved in a separate shapefile (*WSD_ID_L.shp*).

Edition was required on the saved shapefile to remove all polygons representing rivers (Figure 5a). Segmentation was also necessary since large waterbodies having complex shapes were very demanding in terms of software memory and took many hours to process. However, this operation could be subjective considering that waterbodies representing lakes and major rivers are merged together into single polygons when using NHN-CL1 (NHN-Completeness Level One) data. Refer to the NHN Completeness Levels section from the GeoBase web portal for more information about this particular situation.

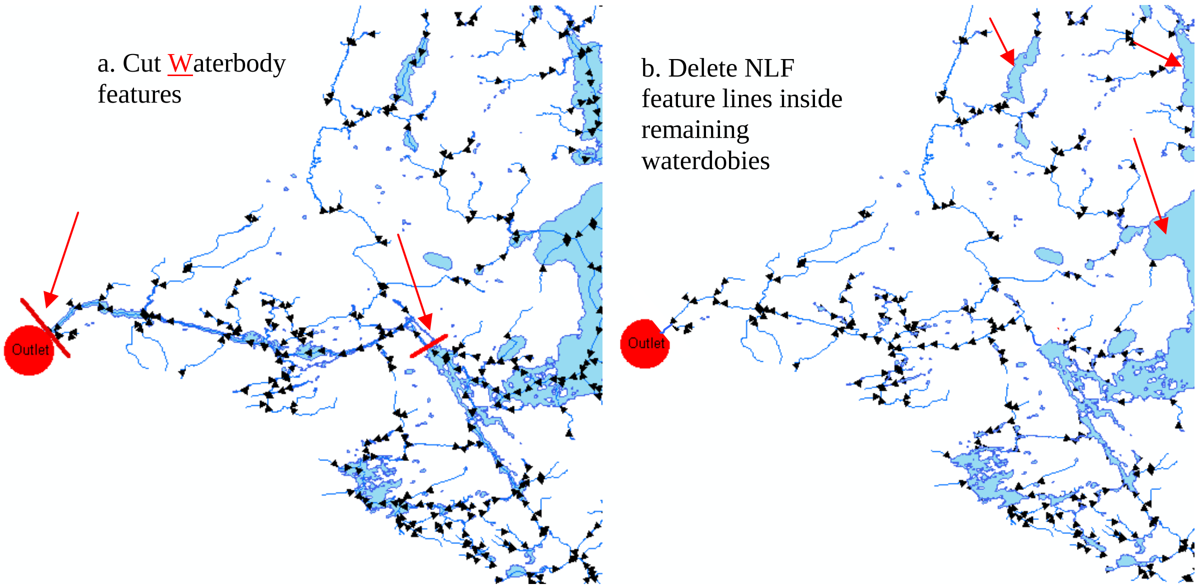


Figure 5. Editing (a) Waterbody polygon features and (b) NLF line features

Still in ArcGIS, the NLF features were split where Waterbody features were cut or segmented. The snapping option of the Edition Toolbar (Figure 6) was enabled to make sure the river NLF node would be connected to a Waterbody feature vertex.

Layer	Vertex	Edge	End
NHN_02JB000_1_0_HN_H	☐	☐	☐
Network02JB_Net_Junctic	☐	☐	☐
CAN_AGR_WSHD_IT	☐	☐	☐
RiverLinesAll_Project	☑	☐	☐
WATERBODY-project	☑	☐	☐
AgMAsk_SLC_Lambert	☐	☐	☐
qc660L	☐	☐	☐
qc660r	☐	☐	☐
Nat_slc_AgrLambert	☐	☐	☐

Figure 6. Snapping options enabled from the Edition toolbar

Finally, all NLF segments falling inside the waterbodies representing lakes were deleted (Figure 5b). This could easily be done using either one of the following queries:

- Using select by attributes create a new selection:
  Select FROM WSD_ID_R.shp WHERE “TYPE” =’2’
  NB "Type" is actually the NLF feature Network Flow Type attribute. A value of "2" is "Inferred" which means the feature is located inside a body of water, excluding liquid waste.

- Using select by location select from the currently selected features
  all features that have their centroid inside WSD_ID_L.shp

- Delete the resulting selected features while within an editing session

iii. Outlet Selection

TI values are valid only when all contributing areas are accounted for during matrix calculations. It should be noted that NHN Work Units or datasets sometimes represent a drainage area that may include many outlets (especially along the coast) and sometimes represent a portion or part of a more important watershed. It was important to have one single file for each single outlet watershed considered in the TI calculation. Coordinate values of each outlet could be identified using ArcGIS or PHYSITEL software capabilities.

C. Format Transformation

i. Geometry Collection

Once editing of NLF river lines and Waterbody polygons was completed, data were merged into a geometry collection[4] of the MAPINFO *.tab* format using the FME 2006 Workbench (Safe Software) (Figure 7).

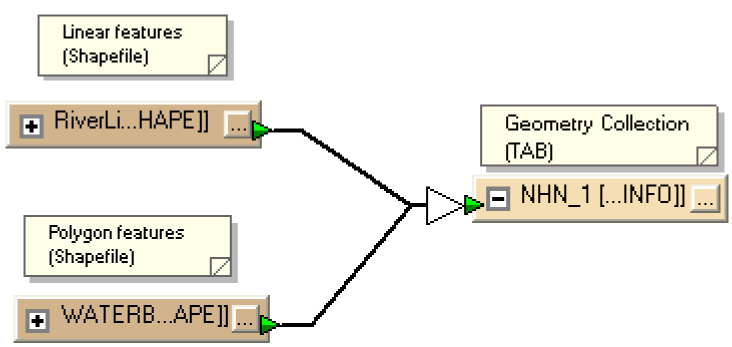


Figure 7. MAPINFO Geometry Collection generation data model

ii. ASCII DEM

Watershed DEMs needed to be converted into a ASCII format in order to be read by PHYSITEL. ArcGIS features provided a conversion tool (Figure 8) for this operation. Creation of Python™ [5] batch programs was facilitated by using ArcGIS COM objects. An example of a batch file is found in Annexe 1.

Figure 8. Raster to ASCII format conversion process

[4] A geometry collection is defined as a collection of different geometries into a single file.

[5] www.python.org

2. TI Matrix Computing

TI calculations were divided into a series of calculations using algorithms from the PHYSITEL software. The first step was to evaluate topology compliance of the Lake-River Network to ensure the creation of a connected drainage system that identified both lakes and rivers flowing to the outlet (Figure 9). The resulting raster data were then used to force the stream network or «*burn*» the DEM, the whole in order to direct the flows in or towards known water channels.

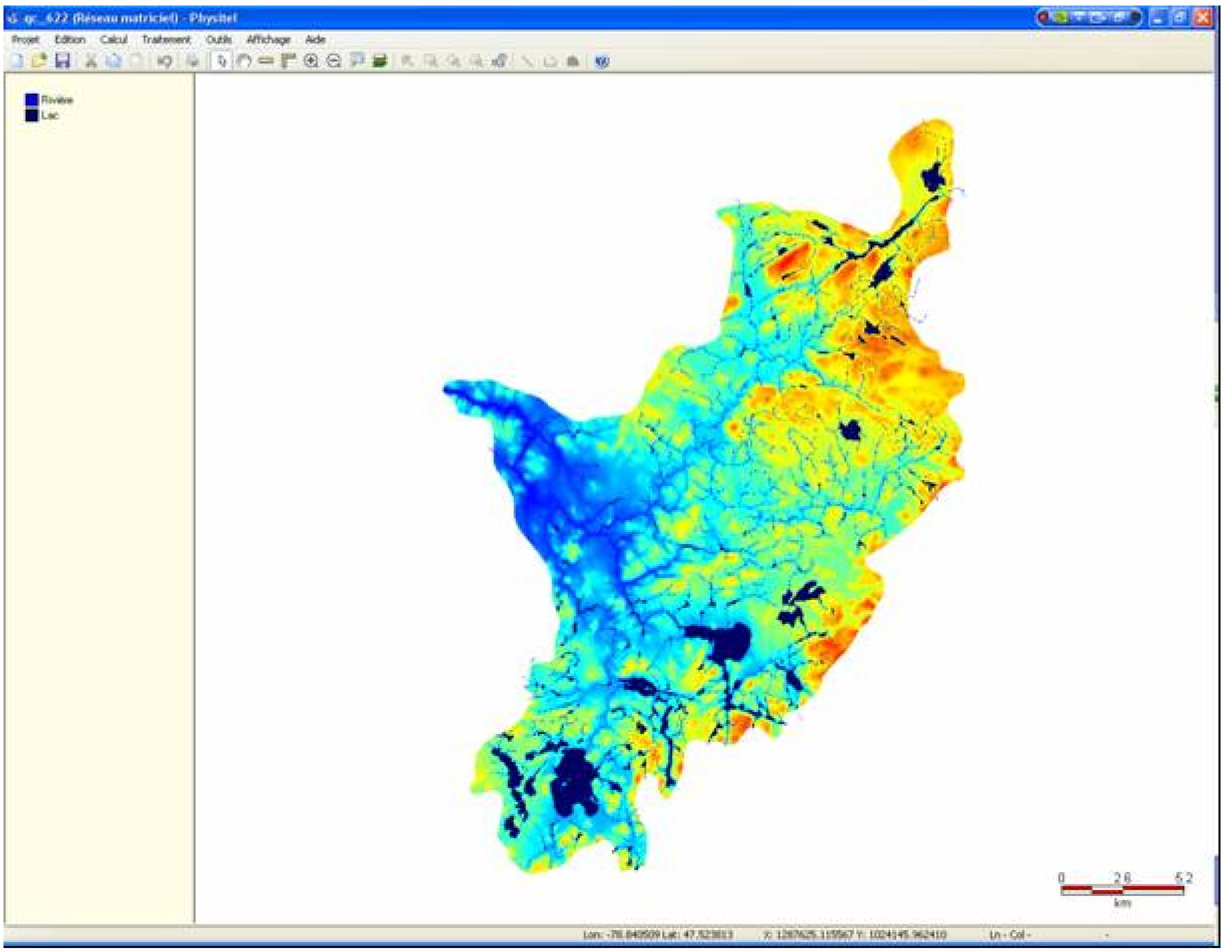


Figure 9. DRLN raster

Local slopes (in percentage) were evaluated from the original DEMs. Flow matrices were created using the D8-*Least Transversal Deviation* algorithm for both surface and DRLN rasters. The flow accumulation was calculated for each pixel. The flow accumulation values for a lake correspond to the number of pixels flowing to the outlet of the lake. For more information regarding the computational procedure of the TI within PHYSITEL, refer to Hentati *et al*. [2005].

TI matrices were then computed using equation 3. An example of the resulting matrix is presented in figure 10.

$$TI = \left(\frac{[Flow\_accumulation] + 1}{\frac{[Slope]}{100} + 1} \right) \quad (3)$$

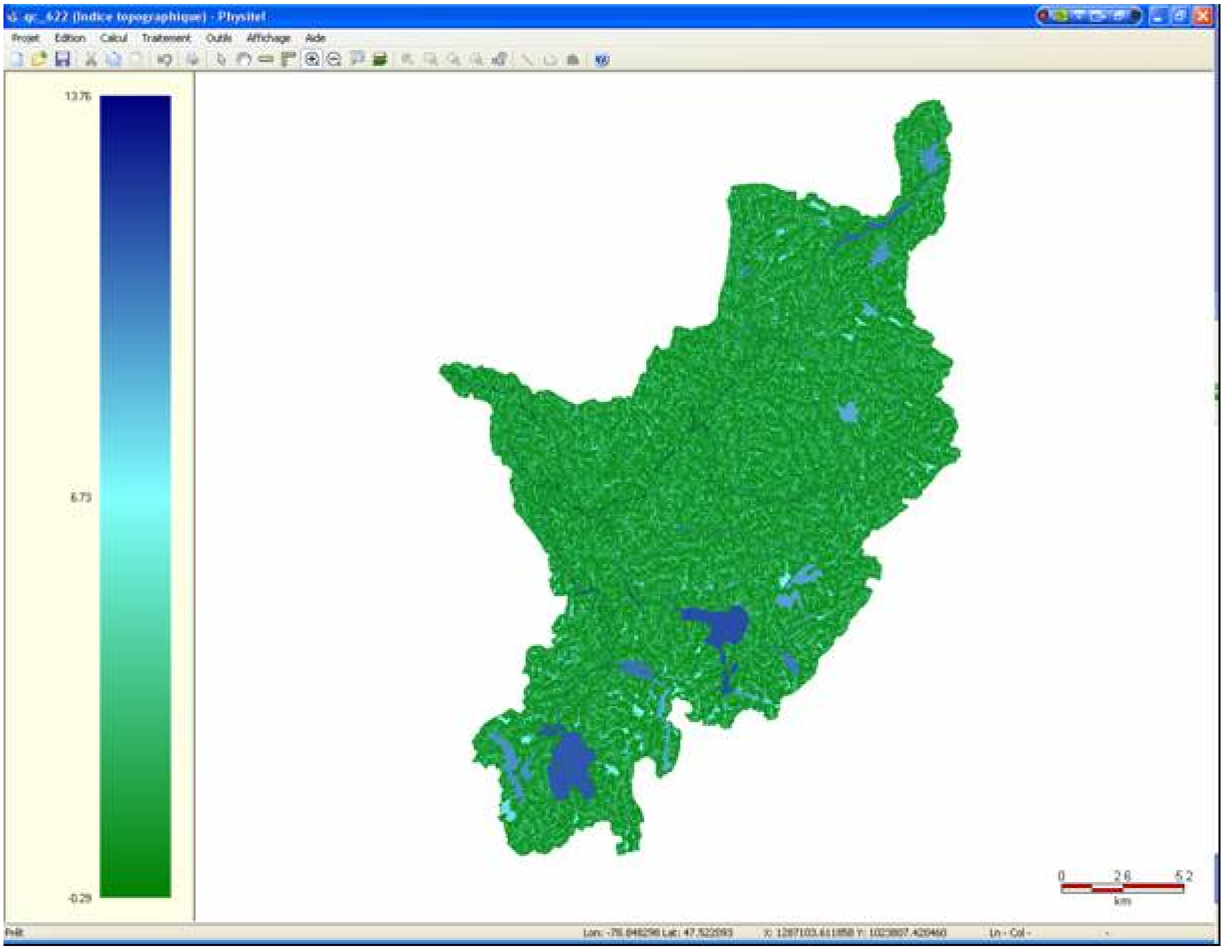


Figure 10. Example of TI matrix

3. TI Upscaling at the Soil Landscape of Canada

Both IROWC-P and IROWC-Col. were developed using the [Soil Landscape of Canada (SLC v3.1.1)](http://sis.agr.gc.ca/cansis/)[6] as the spatial basis for their calculation. Aggregating the data at this scale was necessary to have a constant basis of comparison of TI results. To do so, TI values were determined for all watersheds that had a connection with a SLC polygon. The mean TI value within a SLC polygon was the more appropriate statistical moment to represent the local variability of TI since it could be affected by both low and high extreme values. To this end, a batch program was developed according to the process diagram below (Figure 11).

[6] Soil Landscape of Canada v3.1.1: http://sis.agr.gc.ca/cansis/

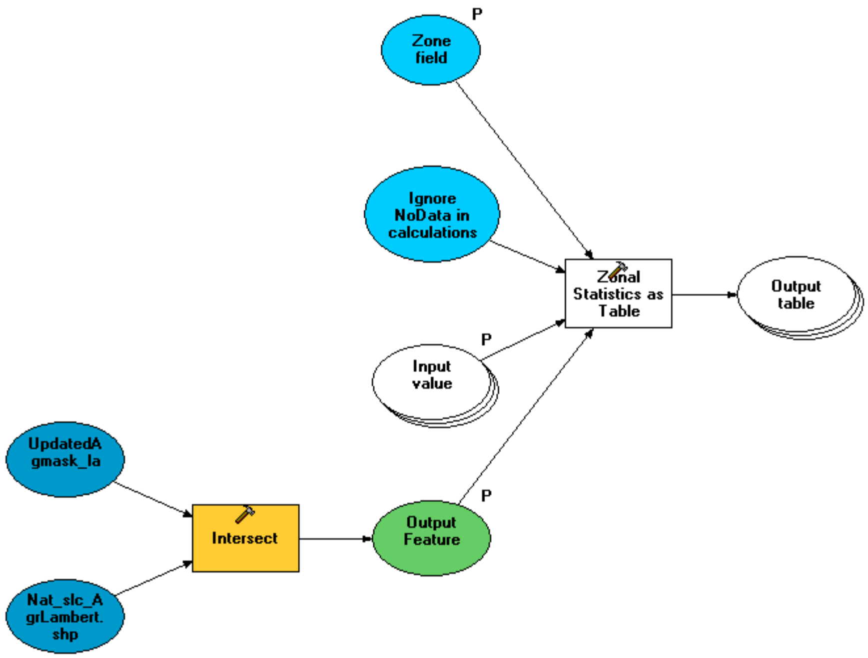


Figure 11. Computing zonal statistics for each SLC polygon

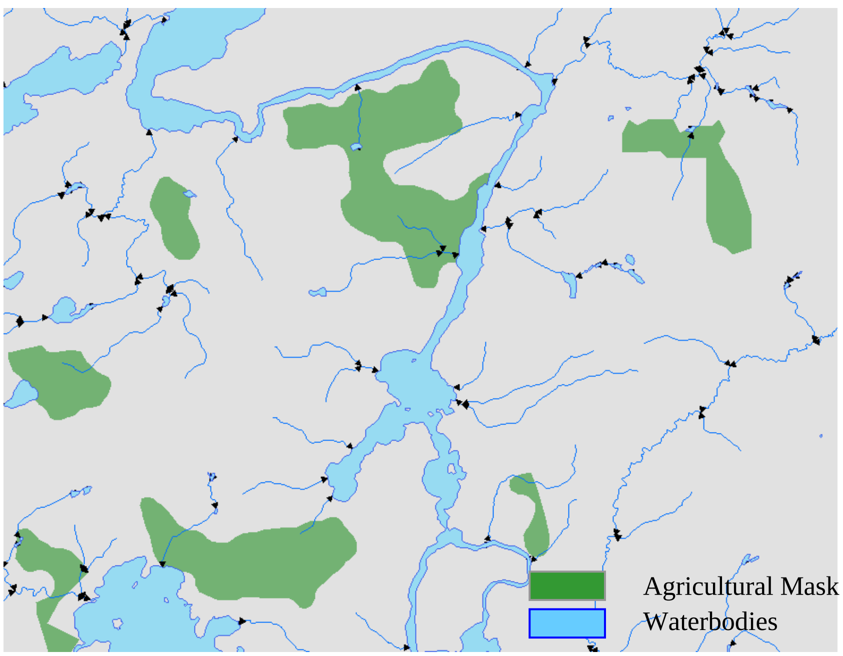


Figure 12. Representation of the agricultural vector mask

In previous diagram (Figure 11)[7], *UpdatedAgMask* represents the spatial distribution of agricultural land in Canada (agricultural vector mask, Figure 12) and *Nat_slc_AgrLambert.shp* is the SLC polygons. The zonal statistics are calculated based on the unique SLC number (zone field in figure 11). All resulting tables needed

[7] The Pyhton batch file is presented in Annexe 2.

to be appended in a final table from which the area-weighted average was calculated. The final TI values were normalized using the distribution of Soil Landscape of Canada TI values to be used as a connectivity factor (Figure 13).

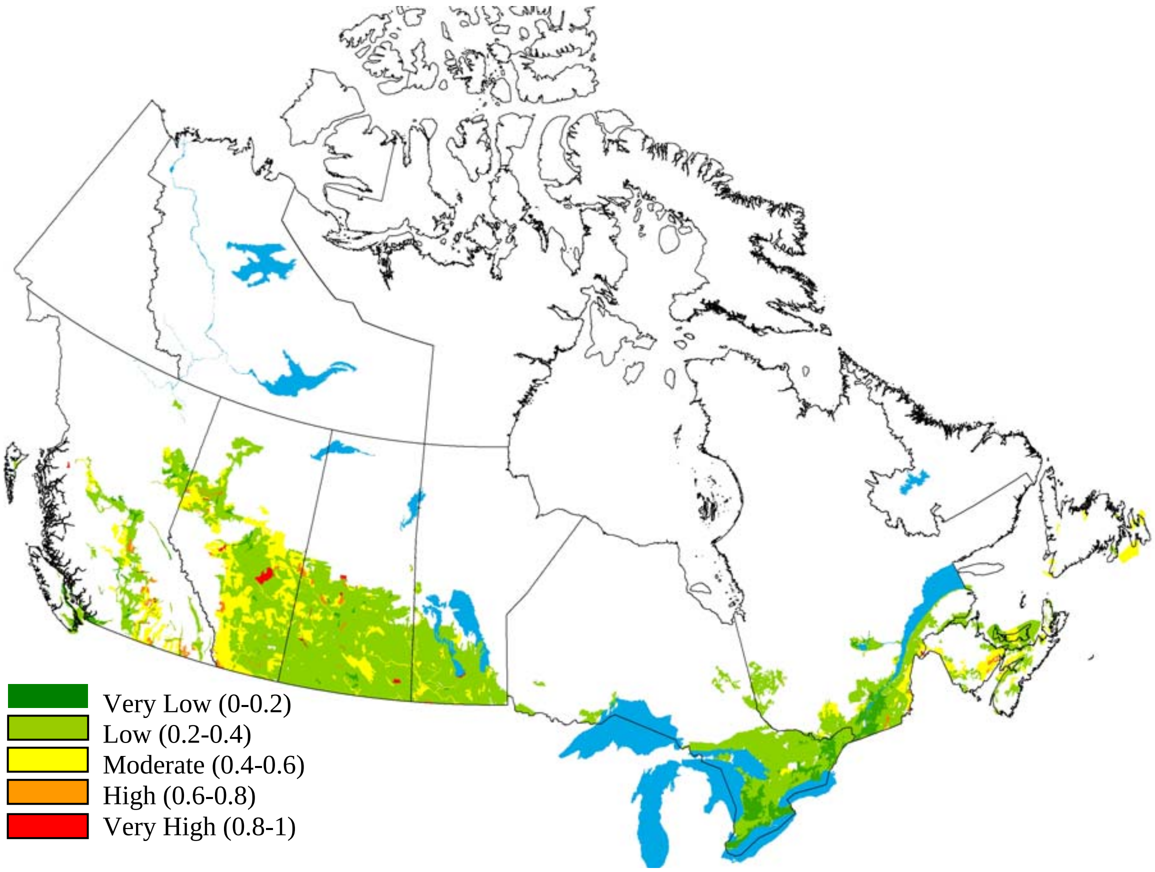


Figure 13. TI connectivity result

The Outcome

Results show that the highest TI values are found in regions where topography tend to be hilly and where agriculture is mostly found in river valleys. The lowest values are found in flat areas with little flow convergence. In the Prairies, localised high TI values are found where agriculture is found close to important waterbodies (lakes or major rivers).

The project was successful since it provided both indicators of risk with a connectivity factor representing the propensity of developing runoff at saturation on agricultural land. The NHN Network Linear Flow and Waterbody features helped us in the generation of well structured DRLN. This contributed to a good characterization of topographic attributes. However, there is still room for improvement. First, the DRLN only works on a well connected drainage network. In the Prairies, a good percentage of the network does not connect with the outlet. These internal drainage or dead drainage areas were forced to reach the outlet creating poorly modeled TI values. To prevent this, all dead drainage areas could have been calculated as single outlet watersheds. Also, a more precise agricultural mask could have probably prevented the data aggregation of pixels considered as permanently saturated and where agriculture is impracticable. Runoff at saturation develops in shallow soils or on soil with impermeable layer close to the surface. Association of TI matrix with soil characteristics using the Soil Landscape of Canada database could help locate and discriminate the regions where those conditions arise.

References

Blazkova S., K.J. P. Beven, Tacheci, A. Kulasova. (2002) Testing of the distributed water predictions of TOPMODEL (allowing for uncertainty in model calibration): the death of TOPMODEL? Water Resources Research, 38(11).

Beven K.J., M.J. Kirkby. (1979). A physically-based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24: 43-69.

Beven K. J. (1997) TOPMODEL: A critique. *Hydrological Processses*, 11: 1069-1085.

Beven K.J., J. Freer. (2001) A dynamic TOPMODEL. Hydrological Processses, 15: 1993-2011.

Garbrecht J., L.W. Martz, (1997). The assignment of drainage direction over flat surfaces in raster digital elevation models. Journal of Hydrology, 193: 204-213.

Hentati, A., A.N. Rousseau. (2005). Computational Procedure of the Topographic Index using PHYSITEL. Rapport de Recherche No R-833. Centre Eau, Terre et Environnement, Institut national de la recherche scientifique, INRS-ETE. Québec, PQ.

Natural Resources Canada, National Hydro Network: Feature Catalogue Distribution Profile, Edition 1.0.1. 06-2008

Orlandini S., G. Moretti, M. Franchini. (2003). Path-based methods for the determination of non dispersive drainage directions in grid-based digital elevation models. *Water Resources Research*, 39(6), 1144.

Turcotte R., J.-P. Fortin, A.N. Rousseau, S. Massicotte, J.-P. Villeneuve. (2001). Determination of the drainage structure of a watershed using a digital elevation model and a digital river and lake network. Journal of Hydrology, 240: 225-242.

Rousseau, A.N., A. Royer, P. Noël, J.-P. Fortin, R. Turcotte, S. Jutras. (2008). PHYSITEL, a specialized GIS for distributed hydrological models. (Paper in preparation for Environmental Modelling & Software)

Royer, A., A.N. Rousseau, J.-P. Fortin, R. Turcotte. (2006). PHYSITEL, un SIG pour la mise en place de modèles hydrologiques. Affiche présentée dans le cadre du Deuxième Symposium Scientifique d'Ouranos sur la Climatologie et adaptation à l'échelle régionale, 2-3 novembre 2006, Montréal, Qc, Canada

Annexe 1. RasterToAscii_Batch.py

```
# ---------------------------------------------------------------------------
# RasterToAscii_Batch.py
# Created on: jeu. mai 10 2007 05:11:10
#   (generated by ArcGIS/ModelBuilder)
# Usage: RasterToAscii_Batch <Input_raster>
# ---------------------------------------------------------------------------

# Import system modules
import sys, string, os, arcgisscripting
try:
# Create the Geoprocessor object
  gp = arcgisscripting.create()

# Load required toolboxes...
  gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Conversion
Tools.tbx")
# Set the workspace. List all of the rasters
  Workspace = sys.argv[1]
  gp.Workspace = Workspace
  Rasters = gp.ListRasters()

    #Reset the enumeration
  Rasters.reset()

    #Get the name of the first raster
  Raster = Rasters.next()
  while Raster:
# Script arguments...
    Input_raster = gp.Workspace + "\\" + Raster

# Local variables...
    Output_ASCII_raster_file = gp.Workspace + "\\" + Raster + ".asc"

# Process: Raster to ASCII...
    gp.RasterToASCII_conversion(Input_raster, Output_ASCII_raster_file)
    Raster=Rasters.next()

except:
  gp.getmessage()
```

Annexe 2. ZonalStats.py

```
# ---------------------------------------------------------------------------
# zonalStats.py
# Created on: jeu. sept. 06 2007 10:04:48
#   (generated by ArcGIS/ModelBuilder)
# ---------------------------------------------------------------------------

# Import system modules
import sys, string, os, win32com.client
try:
# Create the Geoprocessor object
  gp = win32com.client.Dispatch("esriGeoprocessing.GpDispatch.1")
  gp.CheckOutExtension("spatial")

# Load required toolboxes...
  gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Spatial Analyst
Tools.tbx")
# Set the workspace. List all of the rasters
  Workspace = sys.argv[1]
  gp.Workspace = Workspace
  Rasters = gp.ListRasters()
#Reset the enumeration
  Rasters.reset()
#Get the name of the first raster
  Raster = Rasters.next()

  while Raster:

# Local variables...
    Output_table = Workspace + "//"+ Raster + ".dbf"
    Input_raster_or_feature_zone_data = "C:/WorkSpace/AgMask_SLCIntersect.shp"
    Input_value_raster = Workspace + "//"+ Raster
    Zone_field = "SL"
    Ignore_NoData_in_calculations = "true"
# Process: Zonal Statistics as Table...
    try:
      gp.ZonalStatisticsAsTable_sa(Input_raster_or_feature_zone_data, Zone_field,
Input_value_raster, Output_table, Ignore_NoData_in_calculations)
    except:
      print "Does not intersect:", Raster
    Raster=Rasters.next()

except:
    print "end"
```