

Οδηγίες εγκατάστασης του SLEUTH για μοντέλα πρόβλεψης αστικής εξάπλωσης

Κωνσταντίνος Νικολάου

linkedin: <https://uk.linkedin.com/pub/kostas-nikolaou/80/832/428>

mail: konikolaou9@gmail.com

1. Εγκατάσταση του Cygwin

1. Κατεβάστε το Cygwin από το <https://cygwin.com/install.html>
2. Είναι καλύτερο τα αρχεία της εγκατάστασης να τοποθετηθούν απευθείας σε υποφάκελο με όνομα Cygwin στο [C:/](#), και όχι σε κάποιον άλλον υποφάκελο
3. Επιλέξτε τον server <ftp.ntua.gr> ή <ftp.cc.uoc.gr> για Ελλάδα
4. Κατά την εγκατάσταση τοποθετείστε, πέρα από τα ήδη επιλεγμένα, και τα πακέτα “gcc” και “make”

2. Αρχεία του SLEUTH

Κατ' αρχάς, κατεβάζουμε τα αρχεία για το SLEUTH (SLEUTH3.0beta_p01 LINUX released 6/2005) από το

<http://www.ncgia.ucsb.edu/projects/gig/Dnload/download.htm>

Όταν τα κάνουμε εξαγωγή (extract) είναι πιο εύκολο αν τα αποθηκεύσουμε απευθείας σε έναν νέο φάκελο στο [C:/](#) με όνομα SLEUTH.

3. Αρχεία εισαγωγής

Κατ' αρχάς, τα αρχεία που θα χρειαστούμε είναι:

- Slope – Κλίσεις γης
- Landuse – Καλύψεις γης, 2 χρονιές τουλάχιστον
- Excluded – Περιοχές απαγόρευσης της δόμησης
- Urban – Κάλυψη της αστικής γης, 4 χρονιές τουλάχιστον
- Transport – Οδικό δίκτυο, 2 χρονιές τουλάχιστον
- Hillshade – Ανάγλυφο γης

Ετοιμάζουμε τα αρχεία για το SLEUTH. Τα δεδομένα εισαγωγής θα είναι σε μορφή εικόνας, τύπου .gif. Όλες οι εικόνες θα ακολουθούν τρεις βασικούς κανόνες:

- Ίδια όρια
- Ίδιο προβολικό σύστημα
- Ίδια ανάλυση

Όσον αφορά τα επιμέρους δεδομένα, η τελευταία χρονιά του αστικού πρέπει να είναι η ίδια με αυτή των καλύψεων γης. Επίσης, καλό είναι η τελευταία χρονιά του οδικού δικτύου να είναι επίσης η ίδια με αυτή των αστικών και των καλύψεων γης.

Τα δεδομένα πάνω στα οποία δουλεύουμε πρέπει να είναι σε grayscale, με το απόλυτο μηδέν (0) να αντιπροσωπεύει το no values. Η τιμές που δέχονται οι εικόνες ώστε να απεικονίζουν ορθά το θέμα τους και να γίνονται αποδεκτές από το μοντέλο, βασίζονται στο πόσες κατηγορίες ενδέχεται να έχει μια εικόνα. Για παράδειγμα, στις εικόνες με το αστικό, οι τιμές του grayscale είναι 0 (μη αστικό) και 1 (αστικό), όπως επίσης αντίστοιχα στις εικόνες με τις περιοχές αποκλεισμού της δόμησης. Οι εικόνες με τις καλύψεις γης δέχονται τιμές ανάλογα με το είδος στο οποίο ανήκουν. Αυτές οι κατηγορίες δύναται να κατανεμηθούν στο SLEUTH μέσα από τα σενάρια. Εάν κάποιος δεν επιθυμεί να κάνει δικές του κατηγοριοποιήσεις, μπορεί να δημιουργήσει εικόνες καλύψεων γης με βάση τις προϋπάρχουσες που αναγράφονται στα σενάρια, με τις αντίστοιχες τιμές του grayscale να αντιπροσωπεύουν τις κατηγορίες καλύψεων γης:

```
LANDUSE_CLASS= 0, Unclass , UNC , 0X000000
LANDUSE_CLASS= 1, Urban , URB , 0X8b2323 #dark red
LANDUSE_CLASS= 2, Agric , , 0Xffec8b #pale yellow
LANDUSE_CLASS= 3, Range , , 0Xee9a49 #tan
LANDUSE_CLASS= 4, Forest , , 0X006400
LANDUSE_CLASS= 5, Water , EXC , 0X104e8b
LANDUSE_CLASS= 6, Wetland , , 0X483d8b
LANDUSE_CLASS= 7, Barren , , 0Xeec591
```

Αντίστοιχη διαδικασία πραγματοποιείται για την κατηγοριοποίηση του οδικού δικτύου, με το 0 να αντιπροσωπεύει τις τιμές no values και οι τιμές 1, 2 και 3 (ή 25, 50 και 100 αντίστοιχα για όποιον θέλει να χρησιμοποιήσει τις δεύτερες κατηγοριοποιήσεις, χωρίς να αλλάζει τίποτα ουσιαστικά) στην κλίμακα του γκρι να δείχνουν το βάρος του οδικού δικτύου (χαμηλής, μέτριας και υψηλής βαρύτητας οδικό δίκτυο).

Περισσότερες πληροφορίες μπορεί κανείς να βρει στο site του μοντέλου SLEUTH: <http://www.ncgia.ucsb.edu/projects/gig/About/dtInput.htm>

Τα αρχεία που έχουμε ετοιμάσει, τοποθετούνται σε έναν νέο φάκελο στο directory `C:/SLEUTH/input`, με το όνομα του φακέλου στα αγγλικά. Το όνομα του κάθε αρχείου είναι της μορφής (περιοχή).(τύπος_δεδομένων).gif (Για παράδειγμα, volos.excluded.gif ή volos.urban.2000.gif).

4. Εφαρμογή του SLEUTH

Το SLEUTH τρέχει μέσω του cygwin, οπότε ανοίγουμε το τελευταίο με δικαιώματα διαχειριστή (δεξί κλικ --> εκτέλεση ως διαχειριστής – run as administrator)

Οι βασικές εντολές που θα χρειαστεί να γνωρίζουμε είναι:

- “cd”: Change Directory – Αλλαγή φακέλου
- “ls”: List Files – Παρουσιάζει τα αρχεία που εμπεριέχει ο φάκελος στον οποίο βρισκόμαστε
- “cd ..”: Πηγαίνει στον προηγούμενο φάκελο/directory
- “make clean” και “make”: Χρήσιμες εντολές για το compilation/μετάφραση
- “./grow.exe (μία εκ των calibrate, predict, test) (το όνομα του scenario, πχ scenario.volos_predict)”: πχ για την πρόβλεψη η εντολή θα ήταν “./grow.exe predict scenario.demo200_predict”

Για να κατευθυνθούμε π.χ. στον φάκελο `C:/SLEUTH/Whirligif`, πατάμε στην αρχή “cd [C:/](#)”, μετά “cd SLEUTH” και τέλος “cd Whirligif”. Εάν π.χ. από εκεί θέλει

κανείς να γυρίσει πίσω στον φάκελο SLEUTH, τότε δίνει την εντολή “cd ..”

5. Επεξεργασία των αρχείων σεναρίου

Πηγαίνουμε στον φάκελο όπου βρίσκονται τα σενάρια (<C:/SLEUTH/Scenarios>). Από τα σενάρια που υπάρχουν επιλέγουμε τα σενάρια σχετικά με το calibration και το predict, και τα κάνουμε αντιγραφή και επικόλληση στον ίδιο φάκελο, με διαφορετικό όνομα. Το όνομα θα είναι της μορφής scenario (περιοχή)_ (λειτουργία) (π.χ. scenario.volos_predict).

Για να λειτουργήσουν τα σενάρια, τα οποία ουσιαστικά καθοδηγούν το πρόγραμμα προκειμένου να παρουσιάσει τα αποτελέσματα, πρέπει να επεξεργαστούν ώστε να προσαρμοστούν στα δικά μας στοιχεία. Ένα πολύ καλό free πρόγραμμα επεξεργασίας για αρχεία γραμμένα στην γλώσσα C είναι το Netbeans, το οποίο μπορείτε να κατεβάσετε από εδώ:

[http://netbeans.org/downloads/start.html?
platform=windows&lang=en&option=cpp](http://netbeans.org/downloads/start.html?platform=windows&lang=en&option=cpp)

Αφού κατέβει το netbeans και εγκατασταθεί, μπορούμε να το ανοίξουμε προκειμένου να επεξεργαστούμε τα σενάρια. Τα βασικά σημεία που επεξεργαζόμαστε είναι τα εξής:

- Θέση των εικόνων που εισάγουμε (π.χ. ../Input/Volos)
- Θέση των δεδομένων που εξάγονται (π.χ. ../Output/Volos_cal), ανάλογα με το σενάριο
- Τα ονόματα των εικόνων που εισάγουμε
- Monte Carlo Iterations – Επαναλήψεις της διαδικασίας Monte Carlo, αλλάζουν ανάλογα με την διαδικασία
- Coefficients – Οι μεταβλητές οι οποίες παράγονται μετά την διαδικασία της βαθμονόμησης και διαχωρίζονται σε:
 - Calibration – Οι μεταβλητές που ρυθμίζονται για την διαδικασία της βαθμονόμησης

- Prediction – Οι μεταβλητές που ρυθμίζονται για την διαδικασία της πρόβλεψης

6. Compilation – Μετάφραση της γλώσσας προγραμματισμού

Όταν ανοίγει το cygwin, κατευθυνόμαστε στον φάκελο που έχουμε αποθηκεύσει το SLEUTH (στην προκειμένη το [C:/SLEUTH](#)), και πηγαίνουμε στον υποφάκελο Whirlgif. Έπειτα, “καθαρίζουμε” το προηγούμενο cache και κάνουμε το νέο compilation με τις εντολές make clean και αφού τελειώσει με την εντολή make. Ακολουθούμε την ίδια διαδικασία και στον υποφάκελο με directory [C:/SLEUTH/GD](#), καθώς επίσης και στον φάκελο [C:/SLEUTH](#).

Πιο αναλυτικά τα ακριβή βήματα που ακολουθούνται:

Ανοίγουμε το cygwin terminal με δικαιώματα διαχειριστή

```
cd C:/
cd SLEUTH (για να μπει στον φάκελο του SLEUTH)
cd Whirlgif (-----//----- του whirlgif)
make clean (θα γράψει την διαδικασία, κι αφού τελειώσει συνεχίζεις την
διαδικασία)
make
(-----//-----)
cd .. (σε γυρίζει πίσω στον φάκελο του SLEUTH)
cd GD (σε πάει στο GD)
make clean
make
cd ..
make clean
make
```

7. Calibration – Βαθμονόμηση του μοντέλου

Αφού τελειώσουμε με την διαδικασία για το compilation, προχωράμε στην διαδικασία της βαθμονόμησης. Η διαδικασία αυτή χρειάζεται αρκετό χρόνο και θέλει αρκετή προσοχή για την εφαρμογή της.

Για την επίτευξη της βαθμονόμησης, μπορούν να πραγματοποιηθούν δύο διαφορετικές διαδικασίες. Η ρύθμιση του εύρους των προαναφερόμενων μεταβλητών γίνεται μέσω ενός παραγόμενου αρχείου το οποίο μετά από έλεγχο δείχνει τις μεταβολές των παραπάνω μεταβλητών μέσα από στατιστικές. Μία από της μεθόδους για την ρύθμιση των μεταβλητών είναι το OSM (Optimum

SLEUTH Metric) και τα αποτελέσματα που εξάγει μπορεί να χρησιμοποιηθούν για την εισαγωγή του εύρους των μεταβλητών που θα έχουν τις καλύτερες επιδόσεις. Οι βέλτιστοι συντελεστές για το μοντέλο δύναται να βρεθούν και μέσω της διαδικασίας Lee-Salee Metric, για την εύρεση των οποίων ακολουθείται μια πιο χρονοβόρα διαδικασία, αλλά με πιο ακριβή αποτελέσματα. Οι δύο αυτές μέθοδοι μπορούν να καθορίσουν τις βέλτιστες τιμές για τη διαδικασία της πρόβλεψης (prediction best fit) έπειτα από πολλαπλούς ελέγχους.

7.1. Εφαρμογή της μεθόδου βαθμονόμησης Lee-Salee

Η μέθοδος Lee-Salee πραγματοποιεί πολλαπλές βαθμονομήσεις ώστε να μπορέσει να βρει τις βέλτιστες τιμές για την πρόβλεψη. Κατά αυτόν τον τρόπο, τα στοιχεία από τα αποτελέσματα της κάθε βαθμονόμησης χρησίμευσαν για τον καθορισμό του (μειωμένου σε κάθε επανάληψη) εύρους τιμών, ώστε να βρεθούν οι βέλτιστες τιμές για την διαδικασία της πρόβλεψης.

Φάση 1

Πιο συγκεκριμένα, για την εφαρμογή της μεθόδου Lee-Salee ακολουθούνται 3 στάδια: coarse, fine και final calibration (Αναλυτικές οδηγίες στα αγγλικά: <http://www.ncgia.ucsb.edu/projects/gig/Imp/imCalibrate.htm>). Για την εφαρμογή του πρώτου (coarse) σταδίου αλλάζουμε ορισμένα στοιχεία του σεναρίου για την βαθμονόμηση (calibration). Συγκεκριμένα, αυτά που αλλάζουμε είναι τα εξής:

- Monte Carlo Iterations: 4-5
- Όλες οι μεταβλητές (Coefficients) να έχουν τις τιμές:
 - Start: 0
 - Stop: 100
 - Step: 25

Αφού τρέξει το σενάριο με τα παραπάνω στοιχεία (δεδομένου ότι έχουμε κάνει και τις αλλαγές που αναφέρονται στο βήμα 5) πηγαίνουμε στο directory C:/SLEUTH/Output/Volos_cal και ανοίγουμε το αρχείο control_stats.log (κατά προτίμηση με κάποιο πρόγραμμα όπως το LibreOffice calc ή το Microsoft Excel). Έπειτα, κάνουμε sort by Leesalee value, από το μεγαλύτερο στο μικρότερο. Για τις τρεις μεγαλύτερες τιμές τις leesalee (είναι πιθανόν δυο γραμμές να έχουν την ίδια τιμή leesalee) παρατηρούμε τις τιμές για τα πεδία: breed, spread, slope, road gravity (RG) και dispersion (Diff). Από αυτές τις τιμές εξάγουμε την μικρότερη και την μεγαλύτερη, οι οποίες τοποθετούνται ως start και stop τιμές αντίστοιχα για την επόμενη φάση της βαθμονόμησης. Επιπλέον, ως τιμή step για την επόμενη φάση της βαθμονόμησης τοποθετείται ένας αριθμός ο οποίος προκύπτει από την διαφορά μεταξύ της start και stop τιμής, διά ενός αριθμού μεταξύ 4-6. Για παράδειγμα, εάν από τον πίνακα προκύπτει πως για την επόμενη φάση το πεδίο ορισμού είναι 0-20, τότε το step είναι 5. Επίσης, 2 παρατηρήσεις που προκύπτουν είναι:

- Η τιμή 1 στον πίνακα για τις 5 μεταβλητές που αναφέρθηκαν αναγράφεται ως 0 στο πεδίο ορισμού
- Εάν οι τιμές των 5 μεταβλητών του πίνακα είναι μόνο 1, τότε τοποθετείται πεδίο ορισμού 0-20, με βήμα 5, ώστε να ελεγχθεί καλύτερα και στην επόμενη φάση.

Για παράδειγμα, δείτε την εικόνα 1 ή πηγαίετε στην ιστοσελίδα <http://www.ncgia.ucsb.edu/projects/gig/Imp/calCoarseSelect.htm>

Run	Product	Compare	Pop	Edges	Clusters	Cluster Size	sort value		Initial coefficient values										
							Leesalee	Slope	%Urban	Xmean	Ymean	Rad	Fmatch	Diff	Brd	Sprd	Slp	RG	
15	0.08234	0.64331	0.92749	0.92189	0.99241	0.86224	0.34437	0.99988	0.88769	0.91861	0.94748	0.94694	0.6944	1	1	1	75	1	
16	0.08234	0.64331	0.92749	0.92189	0.99241	0.86224	0.34437	0.99988	0.88769	0.91861	0.94748	0.94694	0.6944	1	1	1	75	25	
17	0.08234	0.64331	0.92749	0.92189	0.99241	0.86224	0.34437	0.99988	0.88769	0.91861	0.94748	0.94694	0.6944	1	1	1	75	50	
20	0.00213	0.61402	0.98361	0.9887	0.99985	0.89286	0.33744	0.85379	0.91667	0.97736	0.0221	0.99266	0.7044	1	1	1	100	1	
21	0.00213	0.61402	0.98361	0.9887	0.99985	0.89286	0.33744	0.85379	0.91667	0.97736	0.0221	0.99266	0.7044	1	1	1	100	25	
22	0.00213	0.61402	0.98361	0.9887	0.99985	0.89286	0.33744	0.85379	0.91667	0.97736	0.0221	0.99266	0.7044	1	1	1	100	50	
145	0.02022	0.64854	0.98469	0.98736	0.97959	0.7033	0.31584	0.29837	0.91965	1	0.80372	0.99308	0.6728	1	25	1	100	1	
146	0.02022	0.64854	0.98469	0.98736	0.97959	0.7033	0.31584	0.29837	0.91965	1	0.80372	0.99308	0.6728	1	25	1	100	25	
147	0.02022	0.64854	0.98469	0.98736	0.97959	0.7033	0.31584	0.29837	0.91965	1	0.80372	0.99308	0.6728	1	25	1	100	50	

Coefficient range for fine calibration:

coefficient type	{_START - _STOP, _STEP}
dispersion	{0 - 20, 5}
breed	{0 - 25, 5}
spread	{0 - 20, 5}
slope	{75 - 100, 5}
road gravity	{0 - 50, 10}

Εικόνα 1: Παράδειγμα για την εξαγωγή των τιμών για τις 5 μεταβλητές στην 1η Φάση (Coarse)

Φάση 2

Για την εκτέλεση του 2ου βήματος της βαθμονόμησης, εκτελείται το σενάριο calibration με τις τιμές των μεταβλητών που εξήχθησαν από το 1ο βήμα. Προσέχουμε όμως να έχουμε δημιουργήσει έναν άλλο φάκελο για τα παραγώμενα (output) αρχεία, ώστε να μην υπάρξει πρόβλημα, και αναγράφουμε το όνομα του νέου φακέλου μέσα στο σενάριο για την 2η εκτέλεση του σεναρίου βαθμονόμησης. Επιπλέον, θέτουμε την τιμή Monte Carlo Iterations με αριθμό 7-8.

Αφού εκτελεστεί το σενάριο της βαθμονόμησης για την 2η φάση (fine calibration), ανοίγουμε το αρχείο control_stats.log. Έπειτα, ακολουθείται η ίδια διαδικασία με το 1ο βήμα, προκειμένου να εξαχθούν οι τιμές για την τελευταία φάση της βαθμονόμησης. Μία παρατήρηση που προκύπτει είναι πως εάν οι τιμές των 5 μεταβλητών του πίνακα είναι μόνο 1, τότε τοποθετείται πεδίο ορισμού 0-5, με βήμα 1, ώστε να ελεγχθεί καλύτερα και

στην επόμενη φάση.

(<http://www.ncgia.ucsb.edu/projects/gig/Imp/calFineSelect.htm>)

Φάση 3

Από το παραγόμενο αρχείο control_stats.log, κάνουμε πάλι την ταξινόμηση με βάση το leesalee, από το μεγαλύτερο στο μικρότερο. Έπειτα, επιλέγουμε για το σενάριο μόνο από την πρώτη σειρά τις τιμές για τις 5 μεταβλητές, οι οποίες θα χρησιμοποιηθούν ως start και stop, με βήμα 1, για την τελευταία φάση της βαθμονόμησης. Επίσης, οι τιμές Monte Carlo Iterations ορίζονται σε 8-10. Δεν ξεχνάμε πως, όπως και στα προηγούμενα βήματα, να δημιουργηθεί ένα νέος φάκελος για τα παραγόμενα αρχεία, ο οποίος θα αναγραφθεί στο σενάριο της βαθμονόμησης. Τέλος, στο πεδίο που αναγράφει WRITE_AVG_FILE βάζουμε την τιμή YES., προκειμένου να παραχθούν οι μέσες τιμές (<http://www.ncgia.ucsb.edu/projects/gig/Imp/calFinalSelect.htm>).

Φάση 4

Στο τελευταία βήμα, αφού έχουμε τρέξει για τρίτη φορά το σενάριο της βαθμονόμησης, ανοίγουμε το αρχείο avg.log που παράχθηκε από το προηγούμενο σενάριο. Στην σειρά όπου βρίσκεται η τελευταία ημερομηνία της βαθμονόμησης, βλέπουμε τις τιμές για τις 5 μεταβλητές (diffus, spread, breed, slp_res, rd_grav). Αυτές οι τιμές τοποθετούνται στο πεδίο PREDICTION_BEST_FIT του σεναρίου της πρόβλεψης (εικόνα 2, <http://www.ncgia.ucsb.edu/projects/gig/Imp/calForecastSelect.htm>).

run	year	index	sng	sdg	og	rt	pop	edges	clusters	cl_size	diffus	spread	breed	slp_res	rd_grav	%urban	grw_rate	leesalee	grw_pix
0	1950	1	1.15	0.1	22.27	2.28	993.09	490.46	64.65	15.02	1.2	13.13	3.58	80.95	1.61	4.72	2.59	0.58	25.8
0	1970	2	1.55	0.1	38.09	4.17	1664.3	723.58	102.21	15.88	1.47	16.03	4.37	70.02	2.71	7.49	2.63	0.48	43.91
0	1990	3	1.65	0.14	63.07	3.73	2777.29	1033.16	141.46	19.21	1.79	19.55	5.33	52.38	4.47	11.19	2.46	0.41	68.59

(While we're here, notice that this file contains data showing how the metric values change over time. Several of these metrics are created for the control data and written to the LOG_# file if the LOG_BASE_STATISTICS flag in the scenario file is set to YES.)

Coefficient values used to predict growth:

coefficient type	PREDICTION_*_BEST_FIT
dispersion	2
breed	5
spread	20
slope	52
road gravity	4

Εικόνα 2: Παράδειγμα για την εξαγωγή των τιμών για τις 5 μεταβλητές από το τελευταίο στάδιο της βαθμονόμησης (final) για το στάδιο της πρόβλεψης

8. Prediction – Πρόβλεψη

Ανοίγουμε το σενάριο για την πρόβλεψη και, αφού γίνουν όλες οι αλλαγές που αναφέρονται στο πέμπτο βήμα, τοποθετούνται οι τιμές που προέκυψαν από την τέταρτη φάση του έβδομου βήματος. Η τιμή Monte Carlo Iterations ορίζεται σε 100 και το output σε φάκελο με παράδειγμα ονόματος volos_pre στο directory του output. Τρέχουμε το σενάριο και αναλύουμε τα αποτελέσματα.

(<http://www.ncgia.ucsb.edu/projects/gig/Imp/imPredict.htm>)

*Υπάρχει το ενδεχόμενο το σενάριο που κατεβάζετε από το site του gigaopolis να έχει διαφορετικά πρότυπα (π.χ. Τα ονόματα για τα αρχεία να είναι του τύπου volos.urban2000.gif, αντί του volos.urban.2000.gif, ή ακόμα και τα χρώματα να είναι διαφορετικά). Σε αυτήν την περίπτωση ακολουθήστε τα πρότυπα του σεναρίου που κατεβάσατε.