

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>1</b>  |
| 1.1      | General context . . . . .  | 2         |
| 1.1.1    | Problem description . . . . .  | 2         |
| 1.1.2    | Proposed solution . . . . .  | 3         |
| 1.2      | Outline of the thesis . . . . .  | 5         |
| <b>2</b> | <b>Unsupervised competitive ANN-Based clustering</b>   | <b>7</b>  |
| 2.1      | Competitive learning . . . . .   | 8         |
| 2.2      | Soft competitive learning . . . . .  | 10        |
| 2.2.1    | Self-Organizing Maps (SOMs) . . . . .  | 11        |
| 2.3      | Hierarchical Self-Organizing Structures . . . . .  | 12        |
| 2.4      | Fuzzy representation in Hierarchical Self-Organizing Structures . . . . .                                | 15        |
| 2.4.1    | Fuzzy Kohonen Clustering Networks (FKCN) . . . . .   | 16        |
| 2.5      | Fuzzy Growing Hierarchical Self-Organizing Networks (FGHSON) . . . . .                                   | 17        |
| 2.5.1    | The algorithm . . . . .  | 18        |
| 2.5.2    | Some remarks . . . . .   | 23        |
| 2.5.3    | Tuning parameters in the FGHSON . . . . .  | 25        |
| <b>3</b> | <b>Visualization and projection of geospatial data by means of Self-Organizing Maps Component Planes</b> | <b>27</b> |
| 3.1      | Interpreting clusters in the SOM . . . . .   | 28        |
| 3.1.1    | The Unified Distance Matrix representation (U-matrix) . . . . .  | 29        |
| 3.1.2    | The SOM components planes . . . . .  | 32        |
| 3.1.3    | Correlation Hunting . . . . .  | 33        |
| 3.1.4    | Correlation Hunting by using SOM as a projection algorithm . . . . .                                     | 34        |
| 3.2      | The tree-structured component planes representation . . . . .  | 36        |
| 3.2.1    | The algorithm . . . . .  | 36        |
| 3.3      | Applications in agriculture . . . . .  | 40        |

|          |   |            |
|----------|---|------------|
| 3.3.1    | Improving the correlation hunting in a large quantity of SOM component planes. Classification of agroecological variables related to productivity in the sugar cane culture . . . . .   | 40         |
| 3.3.2    | Analysis of Andean blackberry ( <i>Rubus glaucus</i> ) production models obtained by means of artificial neural networks exploiting information collected by small-scale growers in Colombia and publicly available meteorological data . . . . . | 46         |
| <b>4</b> | <b>Clustering of geospatial data by using soft competitive learning approaches</b>  | <b>56</b>  |
| 4.1      | Unsupervised soft competitive learning algorithms for clustering and projection of geospatial data . . . . .  | 59         |
| 4.1.1    | Self-Organizing Maps (SOMs) . . . . .   | 61         |
| 4.1.2    | Growing Hierarchical Self-Organizing Map (GHSOM) . . . . .  | 65         |
| 4.2      | Fuzzy Growing Hierarchical Self-Organizing Networks (FGHSON) as a tool to explore geospatial data . . . . .   | 68         |
| 4.2.1    | Finding zones with analogous precipitation and air temperature in South America . . . . .   | 69         |
| 4.2.2    | Case study: finding analogous environmental regions in Colombia   | 74         |
| 4.2.3    | Conclusions . . . . .   | 75         |
| <b>5</b> | <b>Finding similar patterns through time in spatio-temporal geospatial data</b>   | <b>79</b>  |
| 5.1      | Related work . . . . .  | 80         |
| 5.2      | Notation and definitions . . . . .  | 81         |
| 5.3      | Finding agroecozones through time . . . . .   | 82         |
| 5.3.1    | Delineation of sugar cane agroecozones . . . . .  | 83         |
| 5.3.2    | Static delineation approach . . . . .   | 85         |
| 5.3.3    | Dynamic delineation approach . . . . .  | 89         |
| 5.4      | Using Spatio-temporal clustering to find similar patterns shifted in time   | 94         |
| 5.4.1    | Using FGHSON to find similar ecoregions shifted in time . . . . .   | 96         |
| 5.4.2    | The dataset . . . . .   | 96         |
| 5.4.3    | Training and results . . . . .  | 98         |
| 5.4.4    | Analysis . . . . .  | 101        |
| 5.5      | Conclusions . . . . .   | 104        |
| <b>6</b> | <b>Conclusions and future work</b>  | <b>106</b> |
| 6.1      | Summary . . . . .   | 106        |
| 6.2      | Original contributions . . . . .  | 107        |
| 6.2.1    | Algorithms contribution for clustering and visualization . . . . .  | 107        |
| 6.2.2    | Test of methodologies for clustering and visualization of georeferenced data . . . . .  | 108        |
| 6.2.3    | Methodology contributions . . . . .   | 108        |

|          |  |            |
|----------|--|------------|
| 6.2.4    | Agroecological knowledge contribution . . . . .  | 109        |
| 6.3      | Contributions for researchers and practitioners . . . . .  | 110        |
| 6.4      | Transferability of my work to other areas . . . . .  | 111        |
| 6.5      | Future work . . . . .  | 112        |
| <b>A</b> | <b>First article</b>   | <b>114</b> |
| A.1      | Introduction . . . . .   | 115        |
| A.2      | Methods . . . . .  | 116        |
| A.2.1    | Self-Organizing Maps . . . . .   | 116        |
| A.2.2    | SOM component planes . . . . .   | 116        |
| A.2.3    | Correlation Hunting . . . . .  | 117        |
| A.2.4    | Distance matrix based clustering of the SOM . . . . .  | 118        |
| A.2.5    | Tree-structured component planes clusters representation . . . . .   | 119        |
| A.2.6    | Problem description . . . . .  | 122        |
| A.2.7    | Classification of agro-ecological variables related with productivity  | 122        |
| A.3      | Conclusion . . . . .   | 124        |
| <b>B</b> | <b>Second article</b>  | <b>126</b> |
| B.1      | Introduction . . . . .   | 127        |
| B.2      | Methods . . . . .  | 130        |
| B.2.1    | Hierarchical Self-Organizing Structures . . . . .  | 130        |
| B.2.2    | Fuzzy Kohonen Clustering Networks . . . . .  | 131        |
| B.2.3    | Fuzzy Growing Hierarchical Self-Organizing Networks . . . . .  | 132        |
| B.3      | Experimental testing . . . . .   | 136        |
| B.3.1    | Iris Data Set . . . . .  | 136        |
| B.3.2    | Toy set . . . . .  | 137        |
| B.3.3    | Tuning the model parameters . . . . .  | 137        |
| B.4      | Conclusion . . . . .   | 144        |
| <b>C</b> | <b>Third article</b>   | <b>145</b> |
| C.1      | Introduction . . . . .   | 146        |
| C.2      | Materials and methods . . . . .  | 149        |
| C.2.1    | Variable selection . . . . .   | 150        |
| C.2.2    | Computational models . . . . .   | 151        |
| C.3      | Results and discussion . . . . .   | 154        |
| C.3.1    | Model performance . . . . .  | 154        |
| C.3.2    | Analysis of the variables relevance . . . . .  | 155        |
| C.3.3    | Visualization of the relations between the variables found as relevant by the sensitivity metric and clusters with similar productivity of Andean blackberry . . . . . | 156        |
| C.3.4    | Component planes and variable dependencies . . . . .   | 156        |

C.4 Conclusions . . . . . 163