Potential for a Managed Aquifer Recharge (MAR) project in the Namoi basin, Australia.

Ignacio Fuentes San Román
Faculty of Sciences
The University of Sydney
2020

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy
This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Ignacio Fuentes
Summary

Over the last two decades, it has become increasingly obvious that human activities have a holistic impact on Earth systems. While in the last century most policies were focused on the exploitation of resources to satisfy human necessities, but also human greed, this century has shown the environmental impacts of such an approach. Most ecosystems are threatened by human activities, a sixth mass extinction event has been recognised, global warming is widely accepted by the scientific community, and natural resources scarcity is commonly discussed in news media and assumed as an ongoing issue that threatens population.

Considering the scenario of water scarcity, this study focuses on water resources at a catchment scale, studying the potential of a managed aquifer recharge (MAR) project to alleviate groundwater resources exploitation and promote groundwater storage in seasons of surface water surplus caused by climatic events.

Costs of a MAR project increase considerably using injection wells, because these not only include all the hydraulic infrastructure required to divert the source water, but also the costs of treatment for the water to be injected, limiting groundwater pollution and clogging risks. Therefore, this thesis is focused on spreading or in-channel modification as alternatives for MAR in the catchment, which ultimately implies the use of streamflow or flood waters as the source of water for the project.

An introduction to the different topics discussed in this thesis is presented in Chapter 1. This addresses global problems including water scarcity, overexploitation of hydric resources, and climate change and its effects on water resources. Additionally, the potential of remote sensing in hydrologic studies was also discussed, and groundwater recharge as a component of the water budget was further analysed. It concludes with a description of MAR and its potential impact as a groundwater alleviation practice.

Since one of the arguments to support MAR projects implies the reduction in evaporation losses, long-term climatic and surface water trends at the catchment scale were firstly evaluated in Chapter 2, which included open water evaporation. Additionally, water usage was also studied as one of the potential drivers of the observed trends.

Chapter 3 of this study addressed the spatiotemporal study of floods in the catchment and its relationship with climatic and hydrologic covariates, which not only allows for an understanding of the spatial patterns of floods in a catchment but also for an assessment of their recurrence, having a potential for flood forecasting.

Nevertheless, the occurrence of floods by itself does not allow for a quantification of available water. As a consequence, in Chapter 4 different algorithms for surface water volume estimation were compared against a proposed methodology, all of
which combine the use of surface reflectance data and digital terrain models (DTM). This study was applied in surface water reservoirs since these have available time series of their storage volumes.

Subsequently, Chapter 5 presents a study of some of the different algorithms evaluated in Chapter 4 in the Namoi catchment, to obtain a spatiotemporal analysis of available water derived from floods in the catchment.

However, the amount of water available for a MAR project is not the only factor in an analysis of its potential application. The nature and characteristics of the medium that might store the water are fundamental aspects to assess its potential. Since there is a huge amount of geoscientific information stored as descriptive data, a model for using this as vectorial information applied specifically for geoscientific use was developed in Chapter 6.

Thereafter, in Chapter 7, the model developed in Chapter 6 was applied to a lithological dataset for New South Wales, which contains lithological descriptions and associated geolocations. Supervised classification and simple interpolation techniques were applied to obtain 3D lithological models. Therefore, this chapter demonstrates the use of natural language processing techniques for the geosciences.

The site suitability for a MAR project was addressed in Chapter 8 through a multi-criteria decision analysis (MCDA) technique, which involved the use of analytical hierarchy process (AHP) and Saaty’s pairwise comparisons on a set of GIS criteria, including surface, aquifer, and underground geospatial information. These allowed a map of suitability to be obtained for the development of MAR projects. From this map, and using a sensitivity analysis, an area of interest (AOI) was selected to allocate the potential MAR project. The natural groundwater recharge in the AOI was estimated from hydrograph data to validate the map.

Finally, Chapter 9 combined all previous chapters into a general discussion. Future lines of work are also enumerated and a final overall conclusion is presented.
Acknowledgments

Firstly, I want to thank my family for all their support, and for all the psychological therapy that I required in this period, which I received candidly from my mother. I also thank my partner for being there, suffering all this experience beside me.

I feel grateful towards those friends that were around in this period, sometimes supporting me, sometimes mocking me to stop thinking too much, and sometimes just sharing the same hate for how this world works.

I thank Willem, my supervisor, and Floris, my co-supervisor, for all the support, suggestions, their constant good mood, and their patience to work with me and with other international students. For sure they have won a place in heaven (if such thing exists).

Finally, since these studies were supported by the national scholarship of the Chilean Government, I thank Becas Chile, and particularly to the Comisión Nacional de Investigación Científica y Tecnológica, CONICYT, for the economic support.
Publications

The following Chapters are published or under revision as research articles in peer review scientific journals or Congress proceedings:

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8
In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Ignacio Fuentes  
2020-01-23

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements included in this document are correct.

Rutger Willem Vervoort  
2020-01-23
## Contents

1. **General Introduction** ........................................................................................................ 1
   1.1. Water overexploitation ................................................................................................. 4
   1.2. Climate change impact on water resources ............................................................... 5
   1.3. Remote sensing potential for hydrologic studies ......................................................... 7
   1.4. Groundwater recharge ............................................................................................... 9
   1.5. Managed aquifer recharge (MAR) ............................................................................... 13
   1.6. Structure of this thesis ............................................................................................... 15
   1.7. References ................................................................................................................ 17

2. **Long-term surface water trends and relationship with open water evaporation losses in the Namoi catchment, Australia** .......... 30
   2.1. Introduction ................................................................................................................ 32
   2.2. Materials and methods............................................................................................. 34
       2.2.1. Study area ........................................................................................................... 34
       2.2.2. Data and preprocessing ..................................................................................... 36
       2.2.3. Surface water detection .................................................................................... 38
       2.2.4. Surface water classes and trend analysis ............................................................ 40
       2.2.5. Evaporation from open water ............................................................................ 41
   2.3. Results ....................................................................................................................... 42
       2.3.1. Water detection .................................................................................................. 42
       2.3.2. Water categories and trends ............................................................................. 45
       2.3.3. Reference evapotranspiration and climatic trends ............................................. 48
       2.3.4. Open water evaporation .................................................................................... 51
       2.3.5. Open water evaporation in the catchment .......................................................... 52
   2.4. Discussion .................................................................................................................. 56
   2.5. Conclusions ............................................................................................................... 60
   2.6. Appendix .................................................................................................................... 61
   2.7. References ................................................................................................................ 62

3. **Spatiotemporal evaluation of inundated areas using MODIS imagery at a catchment scale** ......................................................... 70
   3.1. Introduction ................................................................................................................ 72
   3.2. Materials and methods............................................................................................. 74
       3.2.1. Study area ........................................................................................................... 74
   

viii
3.2.2. Data and preprocessing ............................................................ 74
3.2.3. Water detection methods ........................................................ 77
3.2.4. Performance assessment .......................................................... 79
3.2.5. OWL threshold ........................................................................ 80
3.2.6. Flood dynamics and frequency analysis ...................................... 81
3.3. Results and discussion ................................................................... 82
  3.3.1. General comparison and performance ........................................ 82
  3.3.2. Flood dynamics and relation with covariates ............................... 85
  3.3.3. Frequency analysis .................................................................. 91
  3.3.4. Final ranking of the algorithms .................................................. 94
  3.3.5. General discussion ................................................................... 95
3.4. Conclusions .................................................................................. 96
3.5. Appendix ..................................................................................... 97
3.6. References ................................................................................... 97

4. Comparison of surface water volume estimation methodologies that
couple surface reflectance data and digital terrain models..............103
  4.1. Introduction ............................................................................... 105
  4.2. Materials and Methods ................................................................. 106
    4.2.1. Study areas and data ............................................................. 106
    4.2.2. Water detection ..................................................................... 112
    4.2.3. Water depths ........................................................................ 113
    4.2.4. Covariates and performance of methodologies ......................... 115
  4.3. Results ...................................................................................... 116
    4.3.1. Menindee lakes ..................................................................... 116
    4.3.2. Oklahoma reservoirs .............................................................. 119
    4.3.3. Texas reservoirs .................................................................... 121
  4.4. Discussion .................................................................................. 124
  4.5. Conclusions ................................................................................ 128
  4.6. References ................................................................................. 129

5. Comparing volume estimates from flood detection algorithms for
spatiotemporal analysis of floods ...................................................134
  5.1. Introduction ............................................................................... 136
  5.2. Materials and methods ................................................................. 137
    5.2.1. Study catchment ................................................................. 137
9.1.2. Geologic framework ............................................................... 246
9.1.3. Groundwater models .............................................................. 247
9.1.4. General MAR characteristics .................................................... 248
9.1.5. Regulations and economic aspects related to MAR ................. 249
9.2. Conclusions ........................................................................... 251
9.3. Future work ....................................................................... 252
9.4. References ........................................................................... 253
List of Figures

Figure 1. 1. General schema depicting the different methodologies and products obtained in this thesis. .......................................................... 16

Figure 2. 1. Study area and climatic characteristics. ..................................................35

Figure 2. 2. Landsat images covering the study catchment associated to the tiles of the Landsat satellite constellation. .......................................................... 36

Figure 2. 3. Surface water classification examples. The upper row of images correspond to false colour landsat images (NIR-Red-Green) of Keepit (left) and Splitrock (right) dams, which have been pre-processed. Dry land (yellow) and surface water (turquoise) polygons were manually drawn to feed the random forest classifier. The lower images correspond to surface water classified rasters derived from the upper images. ........................................................................... 39

Figure 2. 4. Surface water recurrence patterns in the Namoi catchment.......... 44

Figure 2. 5. Components of surface water and their evolution in the study period using the RF classifier (a), obtained from the JRC dataset (b) and streamflow recorded in the Namoi river at Goangra (c). .......................................................... 45

Figure 2. 6. Autocorrelation of mean surface water extent at increasing lags. ..... 46

Figure 2. 7. Surface water frequency trends using MK and SS and statistical significance. ........................................................................................................ 47

Figure 2. 8. Time series of the number of on-farm dams and associated extents. 48

Figure 2. 11. Cross correlations between rainfall and surface water extent (blue), and between ET, and surface water extent (red). .............................................. 51

Figure 2. 12. Scatterplot of the PML_V2 evaporation and the ASCE estimated evaporation in four different water bodies evaluated (Keepit, Splitrock and Chaffey dams, and Goran lake) (a) and part of the overall time series of open water evaporation from both datasets at the Keepit reservoir focusing on the period between 2012 and 2015 to highlight the correlation (b)................................. 52

Figure 2. 13. Spatial open water evaporation trends using MK (left panel) and SS (middle panel) in the catchment, and their p-values (right panel). .................. 53

Figure 2. 14. Lumped monthly evaporation in the catchment............................ 54
Figure 2. 15. Seasonal decomposition of lumped evaporation using an additive model. ................................................................. 55

Figure 2. 16. Ratio between evaporation and surface water extents and their trends. ........................................................................................................ 56

Figure 2A. 1. Comparison of surface water in both the JRC dataset and the RF classified images evaluated at Bogabri for three different dates. ......................... 61

Figure 2A. 2. Comparison of surface water in both the JRC dataset and the RF classified images evaluated at Goangra for three different dates. ...................... 61

Figure 2A. 3. Comparison of surface water in both the JRC dataset and the RF classified images evaluated at Wee Waa for three different dates. ..................... 62

Figure 3. 1. Namoi catchment location and spatial distribution of gauge stations. ...................................................................................... 76

Figure 3. 2. Kappa coefficients of the OWL algorithm compared with Landsat surface water detection using different $f_w$ thresholds in MODIS Terra images for different dates................................................................. 81

Figure 3. 3. Boxplot of the distribution of inundated areas calculated using the different approaches on a daily basis. ..................................................... 82

Figure 3. 4. Time series of flood extents (left axis) obtained through the different algorithms used and daily discharges from the Namoi River at Walgett (right axis) from July 2000 to June 2001 (a) and July 2011 to June 2012 (b). ......................... 86

Figure 3. 5. Scatterplots between daily discharge and inundated areas obtained from the different algorithms at two gauge stations located in the Namoi River, at Walgett (a), and at Bugilbone (b), and their correlation coefficients. The dashed line is the linear fit for each algorithm. ........................................................................ 87

Figure 3. 6. Maximum cross-correlation coefficients between daily discharges and daily inundation areas obtained from the ISL algorithm for different gauge stations (above) and their corresponding time lags (below). ......................................................... 88

Figure 3. 7. Cross-correlation between daily lumped rainfall over the catchment and flooded areas for the different algorithms. .............................................. 90

Figure 3. 8. Determination coefficients ($r^2$) from second-degree polynomial regressions for inundated area extents at increasing accumulated rainfall lags for daily MODIS images (a), and eight-day composite images (b). ....................... 91
Figure 3. 9. Occurrence probabilities obtained from the recurrence layer of the GSW product (above), the ISL algorithm (middle), and the OWL algorithm (below). ... 92

Figure 3. 10. (a) Frequency curves obtained using daily and composite MODIS imagery and through the GSW product. (b) Flood frequency curves obtained from the discharge time series in the Namoi River. ................................................. 93

Figure 3.A 1. Maps of water detection for three different dates with the results of Landsat 5 and the different MODIS-based algorithms, for a section in the downstream part of the catchment. Note that the Landsat image is missing some of the areas, while some of the approaches use different time windows. As a result, there are slight differences in the total area displayed between methods. ..... 97

Figure 4. 1. Location of Menindee lakes in NSW, Australia. .............................................. 107

Figure 4. 2. Location of Oklahoma water reservoirs in the United States. ........ 108

Figure 4. 3. Texas water reservoirs selected. ................................................................. 109

Figure 4. 4. Time series of volume storages for the Menindee lakes and LiDAR elevations. One GL is equivalent to 10^6 m^3. ................................................................. 110

Figure 4. 5. Terrain elevation of Oklahoma lakes. .................................................. 110

Figure 4. 6. Terrain elevation of Texas lakes. ...................................................... 111

Figure 4. 7. Two examples of areas where polygons classified as water (blue polygons) and dry (yellow polygons) were drawn to feed the classification and regression tree analysis (CART) classifier with their corresponding classifications. The background images where the polygons were drawn correspond to color infrared Landsat images. ................................................................. 113

Figure 4. 8. Schema of inverse distance weighting interpolation (IDW) methodology for water volume estimation. ................................................................. 115

Figure 4. 9. Water volume estimates (a), residuals (b) and area-volume relationship (c) obtained using the inverse distance weighting interpolation methodology on the Cawndilla, Menindee and Pamamaroo lakes. The cumulative frequency of slopes in the reservoirs is also presented (d). ................................................................. 118

Figure 4. 10. Water volume estimations (a), their residuals (b), and the area-volume relationship (c) obtained using the inverse weighted distance interpolation methodology on the Oklahoma reservoirs. The cumulative frequency of slopes in the reservoirs is also presented (d). ................................................................. 120

Figure 4. 11. Water volume estimations (a), their residuals (b), and the area-volume relationship (c) obtained using the inverse distance weighting interpolation
methodology on the Texas reservoirs. The cumulative frequency of slopes in the reservoirs are also presented (d). ............................................................... 122

Figure 4. 12. Hubbard Creek lake satellite image taken on 15 February 1991 and the different water depth maps obtained using the different methodologies. ........... 123

Figure 4. 13. Effect of pixel size on the volume estimation errors for the Cawndilla (a) and the Hubbard (b) reservoirs. The slope differences between reservoirs are also presented (c). .................................................................................... 124

Figure 4. 14. Area errors produced by pixel representation (dashed lines) of actual terrain elevation (continuous lines) for steady (a) and curved (b) slopes. Blue and red areas represent area gains and losses, respectively. ......................... 126

Figure 5. 1. Time series of inundation volumes estimated and daily discharges at Walgett. 140

Figure 5. 2. Cross correlation between inundated volumes and daily rainfall. .... 140

Figure 5. 3. Scatterplots of monthly accumulated rainfall and monthly accumulated inundation volumes. .............................................................................. 141

Figure 5. 4. Scatterplots of inundated volumes against daily discharge volumes in the Namoi river at Bugilbone (above) and Walgett gauge stations (below). ...... 142

Figure 5. 5. Maximum cross correlation coefficients between daily discharge at different locations in the catchment and inundation volumes (left below) obtained within different time lags (left above), and occurrence probability maps using daily (MOD09GA; right above) and 8 days composite (MOD09A1; right below) imagery. .............................................................................................................. 143

Figure 5. 6. Proportion of water losses other than stream discharge obtained through the deviation of regression slopes from a 1:1 line. ................................. 144

Figure 5. 7. Flood frequency curves obtained from the temporal analysis of flood volume estimations (left), and from daily discharges in the Namoi river at Bugilbone and at Walgett (right)................................................................................ 145

Figure 6. 1. Example of two encodings of the phrase “red sticky clay”: numerical encoding and one-hot encoding........................................................................... 152

Figure 6. 2. Examples of two-dimensional PCA projection of selected word embeddings using a general domain model. The figure illustrates the country–capital relationship learned by the model. Also notice that the model learned about the geographic relationship between the places. Example adapted from Mikolov et al. (2013b). ....................................................................................................... 153
Figure 6. 3. Co-occurrence probability matrix of soil orders (USDA) and selected words................................................................. 158

Figure 6. 4. Overall performance of the embeddings versus number of tokens used to construct the co-occurrence matrix. The improvement limit is around 300 million tokens. For future comparisons, this limit corresponds to approximately 280,000 articles, 22.5 million sentences and 700,000 unique tokens.......................... 160

Figure 6. 5. Two-dimensional PCA projection of selected words. Simple syntactic relationship between particle fraction sizes and rocks (a) and advanced semantic relationship between rocks and rock types (b).............................................. 161

Figure 6. 6. Two-dimensional PCA projection of selected categorisations. Clusters representing soil types from different soil classification systems (a) and a different aggregation level where the same soil types are grouped as a single cluster when compared with rocks (b)........................................................... 162

Figure 6. 7. Interpolated embedding in a two-dimensional PCA projection showing a size gradient (a) with clay < silt < sand < gravel < cobble < boulder; and a gradient of the metamorphism grade (b) with slate < phyllite < schist < gneiss < migmatite. Red and blue dots represent selected words ("clay" and "boulder", and "slate" and "migmatite") and black dots represent the closest word (cosine similarity) to the interpolated embeddings.......................................................... 163

Figure 6. 8. Convex-hulls of great group embeddings at the order level (soil taxonomy). Great group embeddings were obtained after averaging the embeddings of all the words in the descriptions of the profiles belonging to each great group. The convex-hulls were estimated from the two first principal components of the great group embeddings....................................................... 164

Figure 7. 1. Study area and lithologic framework depicting the different orogens that are the main source of sediments filling the sedimentary basins in NSW. Different colours represent the different surface lithologies. ............................................. 176

Figure 7. 2. Boreholes distribution in NSW. .................................................. 177

Figure 7. 3. Frequency of the descriptions length in the dataset used. ............ 178

Figure 7. 4. Example of semantic relationships between types of rocks and lithologies (left) and between mineral groups and minerals (right) in a two dimensional PCA projection obtained by using a domain-specific GloVe model for geosciences (Padarian and Fuentes, 2019). ................................................................. 179

Figure 7. 5. Set of interpolation tokens obtained between two extreme terms showing scales of particle size (left) and metamorphic grade (right; Padarian and Fuentes, 2019).................................................................. 180
Figure 7. 6. Bore density map and selected of areas of interest (AOIs). The zoomed satellite images correspond to false colour (bands 3-4-1) Landsat 5 images. 181

Figure 7. 7. Architecture of the MLP neural network used for the lithological classification. 184

Figure 7. 8. Training size and its effect on the classification accuracies for the training and validation datasets and on the classifier training time. 188

Figure 7. 9. Mean normalised Shannon entropy per lithological class (left panel) and the percentage of each lithological class (right panel) in the different study areas. 189

Figure 7. 10. Interpolation accuracies in depth (left) and number of points used in the interpolation (right). 191

Figure 7. 11. Moree 3D lithological maps obtained using a NN interpolation (left) and a triangulation of embeddings (right). 192

Figure 7. 12. Coleambally 3D lithological maps obtained using a NN interpolation (left) and a triangulation of embeddings (right). 193

Figure 7. 13. Uncertainty mapping for the Moree AOI using the NN interpolation (above) and the triangulation of embeddings (below). Entropy scales were normalised to the range of values obtained. 194

Figure 7. 14. Uncertainty mapping for the Coleambally AOI using a NN interpolation (above) and a triangulation of embeddings (below). 195

Figure 7.A 1. Mean and standard deviation of confusion indices with depth for Moree (left) and Coleambally (right). 198

Figure 7.A 2. Confusion index histograms for the Moree (left) and Coleambally (right) interest areas until 100 m depth. 199

Figure 8. 1. Namoi catchment location in Australia. The map displays the main stream network, main cities and major roads. 210

Figure 8. 2. Surface hydrogeologic units in the Namoi catchment. 211

Figure 8. 3. Flowchart of the methodology used in the study. 218

Figure 8. 4. Reclassified criteria maps used for the MCDA. 219

Figure 8. 5. Suitability classes map obtained by coupling the reclassified criteria maps and the AHP technique. High values indicate high suitability. 221

Figure 8. 6. Suitability maps for the different weighting schemas used in the sensitivity analysis. 222
Figure 8. 7. Selection of an area of interest (AOI) for the emplacement of a MAR project, which was located in areas of high and very high suitability. The left image corresponds to the suitability classes’ map, while the zoomed in image in the right corresponds to a false color (bands 4-3-2) Landsat 5 image of the AOI. A gauging station at Mollee (Number 419039; light green point) and different monitoring wells (yellow points) were selected for the validation and available water assessments.

Figure 8. 8. Different hydrographs from monitoring wells located in the AOI.

Figure 8. 9. Estimation of recharge through the water table fluctuation method in the monitoring well 25107. From a hydrograph (left upper panel) the delineation of a master recession curve ($R^2 = 0.97$) was carried out (right upper panel), which leads to a projection of groundwater head changes (left lower panel) and a subsequent estimation of recharges (right lower panel).

Figure 8. 10. Mean recharge estimates for different labeled monitoring wells located in the AOI.

Figure 8. 11. Correlation coefficients between recharge estimates and rainfall (left), and between annually aggregated recharge estimates and mean annual discharge at the Namoi River (right).

Figure 8. 12. Rating (left) and flood frequency (right) curves at the Mollee gauging station selected. The flood frequency curve presents both, a logarithmic scale in both axis and a logarithmic regression curve (dashed line).

Figure 8. 13. Cross section profile of the hydrogeological setting derived from boreholes at Mollee.

Figure 9. 1. Main schema summarising the different products obtained in this thesis, their methodologies and how these connect with future works for the development of a MAR project.

Figure 9. 2. Lithological 3D model of the AOI selected in Chapter 8 in the Namoi catchment obtained using a triangulation of embeddings.
List of Tables

Table 2. 1. Landsat tiles and images included in the study catchment. ....................... 37
Table 2. 2. Images and characteristics used as variables in the random forest classifier. ................................................................................................................................. 38
Table 2. 3. Surface water detection performance. .......................................................... 43
Table 3. 1. Sources of information used in the study. ...................................................... 75
Table 3. 2. Different surface water detection algorithms used in conjunction with MODIS imagery. ........................................................................................................ 77
Table 3. 3. Evaluation of accuracy between Landsat water detection and different MODIS-based water detection algorithms. ......................................................... 83
Table 3. 4. Qualitative ranking of MODIS-based water detection algorithms. .......... 94
Table 4. 1. Design characteristics of the studied reservoirs (Solis et al., 2010; Solis et al., 2012a; Solis et al., 2012b; Leber et al., 2018; Shivers, 2016). ................................. 112
Table 4. 2. Performance of the different methodologies used to estimate water volumes in the Menindee lakes. ............................................................ 117
Table 4. 3. Performance of the different methodologies used to estimate water volumes in Oklahoma reservoirs. ................................................................. 119
Table 4. 4. Performance of the different methodologies used to estimate water volumes in Texas reservoirs. ................................................................. 121
Table 6. 1. Search terms used to retrieve full-text articles from Elsevier ScienceDirect APIs ........................................................................................................... 154
Table 6. 2. Evaluation scores for each task for our domain-specific (GeoVec) and general domain embeddings (Stanford). For the analogy task, top-1, 3, 5, and 10 represents the accuracy if the expected word was within the first 1, 3, 5, or 10 words returned by the model. For the relatedness task, the score represents the absolute value of the Pearson correlation (mean of the three human subjects). For the categorisation task, the score represents the mean value of 50 v-measure scores. The possible range of all scores is zero to one, where higher is better. 159
Table 7. 1. Aggregation of lithologies into Major classes. .............................................. 182
Table 7. 2. Confusion matrix generated comparing the semi manual and the automated classification using description embeddings. ................................. 186
Table 7. 3. Classification performance of MLP neural network for the sentence embeddings obtained by averaging the word embeddings of the descriptions. 187

Table 7. 4. Mean performance metrics at Moree up to 60 m depth, and in Coleambally up to 80 m depth. 190

Table 8. 1. Criteria and rating used in the site suitability study. 213

Table 8. 2. Different weighting schemas for the groups of variables (Table 8.1) used in the sensitivity analysis. 216

Table 8. 3. Decision hierarchy of MCDA. 220

Table 8. 4. Suitability areas for the different weighting schemas used. 223

Table 8. A 1. Consolidated pairwise comparison matrix obtained through the AHP technique. Values close to 1 mean variables have equal importance, values smaller or greater than 1 indicate smaller or greater importance of the row criteria compared to the column criteria. 234
Chapter 1

General Introduction

Human activities are directly and indirectly affecting natural fluxes through Earth systems at different scales (Vitousek et al., 2017). These have modified the landscapes at a big scale and at unprecedented rates (Smith and Zeder, 2013). Even though specific activities can have immediate and tangible effects on the surroundings, indirect long-term effects are usually underrated, yet they also play a crucial role in the global functioning of Earth systems (Liu et al., 2007). Not taking into account that all activities have local and global impacts in an interconnected global system is one of the causes of current global issues leading to global climate change, as well as ocean warming and acidification, a sixth large-scale mass extinction in progress, depletion of natural resources, among many others (Steffen et al., 2007).

The XIX and XX centuries were characterised by science promotion and the development of technologies for resource exploitation. It is clear that the satisfaction of human needs was one of the main drivers for such development. Even though some thinkers were starting to debate resource limits, the general perspective was actually not concerned with the limits of resources or their renewal rates (Hubbert, 1956). In previous centuries the population was strongly limited by a lack of medical advances and therefore, per capita resources were not considered an issue, since mortality rates were high (Steffen et al., 2007; Krausmann et al., 2009). Therefore, most regulations and policies were developed under a scenario of resource abundance and in order to foster economic growth and development (Papyrakis and Gerlagh, 2007), sometimes leading to contradictory results (Polterovich et al., 2010).

However, since a strong drop in mortality rates has led to a progressive increase in population, at the end of the XX century it was clear that human exploitation of natural systems had an impact on the surrounding ecosystems, and some studies were starting to take into account the interconnection between ecosystems at a planetary scale (Western, 2001; Power and Chapin, 2010).

The XXI century presents a global scenario which is completely different. All the consequences of “growth” at unsustainable exploitation and production rates are evident and pose a threat to humans and ecosystems (Steffen et al., 2011). The current scenario is more a scenario of global interconnectedness, with commons and resources being translocated between continents, and where the main premise is
scarcity (Pereira et al., 2009). However, most policies and regulations have been inherited from past centuries were conditions were completely different, and the change in these policies is being strongly withstood by factual powers (Napolitano, 2012).

While several studies (Obi, 2000; Tyagi et al., 2014; Singh et al., 2017) address the current environmental issues, at the global scale, as a consequence of world population, FAO (2011) has stated that, at least in agriculture, one third of the food produced to feed humans is lost. This means that all the resources used in its production are wasted. Additionally, agriculture has been pointed out as the main driver of Earth systems exceeding planetary boundaries (Campbell et al., 2017). Similarly, Thoth and Szigeti (2016) identified consumption patterns instead of population as the main driver for economic growth and land degradation. Clearly this implies that population is not the main driver of resource scarcity, but rather the reason why and how food is produced and distributed, pointing in the end to socioeconomic aspects of our societies.

Since production within a free market economy is not intended for human need satisfaction, but for the increase of profit, finite natural resources do not match the marxian theoretically infinite capital accumulation, and therefore a metabolic rift would take place (Foster, 1999). Even though the principle of infinite capital accumulation has been contested (Piketty, 2014), the main objective of production is still more related to economic gains than based on human welfare and ecosystem preservation. However, as pointed out by Hardin (1968), it must be taken into account that it is not possible to increase both, human population and welfare in a planet of limited resources.

A narrow view applied to the understanding of natural systems has been a result of a lack of knowledge and computation capacities of previous centuries. The lack of knowledge, considered as one of the causes of current environmental issues (Western, 2001), has been gradually changing into a more holistic perspective, but this is being confronted by socioeconomic pressures. For instance, the current world population already poses several difficulties for adjusting to the global climatic and resource trends, and it is not expected to stabilize in the current century (Gerland et al., 2014).

Food production is one of the main drivers for resource consumption (Campbell et al., 2017). Agriculture requires a chain of products destined to maintain crop yields or feed livestock. Of these, irrigated crops also need to secure water. This implies the building of hydraulic infrastructure necessary to store and move large amounts of water from different sources (Cosgrove and Loucks, 2015). Additionally, crops require land-use changes, which have been implemented extensively. These changes ultimately affect infiltration, soil water retention capacity, and runoff, whose
effects are aggregated when taken into account at the regional or higher scales, resulting in a modification of natural stream flow patterns (Castello and Macedo, 2015). However, other activities, such as hydroelectric power generation, also lead to a modification of landscape, which results in a modification of natural hydrologic flow patterns and even to local climate changes (Pringle et al., 2000).

Therefore, hydrologic changes are not only a consequence of direct water withdrawals but a result of different human activities modifying the land surface and emitting associated waste, which translate into ecosystem pollution, increase in greenhouse gases concentrations, and natural resources cycling disruption (Steffen et al., 2011; Cosgrove and Loucks, 2015).

In addition, several studies have reported the depletion of groundwater resources due to unsustainable exploitation rates (Döll et al., 2014; Gleeson et al., 2012). This also has several negative implications, such as decreasing water availability, which affects ecosystems that rely on groundwater, but also leads to aquifer compaction, resulting in a loss of transmission capacities (Domenico and Schwartz, 1998), among others. All these accumulate to the aforementioned hydrologic changes.

In order to quantify the sustainability of resources, ecological footprints have been conceptualized, which involves the consumption and use of natural resources (Geng et al., 2014; Hoekstra and Wiedmann, 2014). These footprints allow comparison of consumption versus production rate, or recovery rate, of natural resources in order to analyse the sustainability of consumption habits.

In the case of water resources, the water footprint has been formulated, which has allowed water usage to be separated into different components, depending on its stock and availability (Hoekstra and Mekonnen, 2012). Thus, blue water has been defined as water from surface or groundwater sources used for consumption. Green water corresponds to the water that infiltrates and is stored in the soil, which is subsequently lost by evapotranspiration, and therefore sustains vegetation growth. Gray water is the water required to assimilate and dilute contaminants up to a point that it meets standard quality requirements (Ridoutt and Pfister, 2010). All these three components may be studied in isolation or jointly in order to have a thorough understanding of water usage and their impacts on the environment.

However, other schemas can be applied to understand the water resources dynamics. Probably the most useful refers to the different components of the water budget (Trenberth et al., 2007). Under this schema, the different fractions of water are taken into account in such a way that a balance between inputs and outputs in the analysed system must be reached (Scanlon et al., 2002). Using this methodology, the level of abstraction can be modified by the users in order to study the different components, depending on the depth and aims of the study.
To get a thorough understanding of modifications in the hydrologic cycle, and in order to evaluate the impacts of water usage and climatic trends on water availability, different components in the water budget must be studied (Scanlon et al., 2002). These studies may ultimately lead to integrated management plans which can be implemented taking all these changes into account in order to preclude further deterioration in water resources, associated ecosystems, and consequently, in additional socioeconomic affectation.

Nevertheless, since water trends are a result of a combination of processes, water consumption must firstly be discussed in order to understand the effects of direct withdrawals on the global water resources and their associated ecosystems.

### 1.1. Water overexploitation

World water consumption trends are being gradually quantified. Currently, the different usages of water can be broadly subdivided into agricultural, municipal, and industrial. Agricultural water withdrawals represent over 70% of global water usage, while municipal and industrial withdrawals represent around 10 and 20% of water withdrawals, respectively (FAO, 2016). It is important to point out that agricultural withdrawals also represent approximately 90% of total water consumption, as part of total usage (Davis et al., 2015).

Therefore, food production must be considered as the main cause of water consumption. Additionally, agriculture strongly relies on climate conditions. Therefore, global changes in climatic variables impact directly on crop production. Based on this relationship, several studies are taking into account the food-water nexus in order to analyse the current scenario and point out different management practices to cope with the negative impacts of climate change (Biggs et al., 2015; Smajgl et al., 2016; Scanlon et al., 2017).

Among the global terrestrial sources of water, strong negative trends in their volumes can be detected in polar regions associated with ice-sheet and glacier losses, but these are mainly related to global warming. However, there are also areas in Antarctica and Greenland where a positive trend is observed, where snow accumulates (Rodell et al., 2018). In other regions, different factors are triggering changes in terrestrial water volumes. For instance, mid latitudes are experiencing a depletion of non-frozen water sources, which according to the IPCC models is related to a decrease in rainfall.

In addition to climate, withdrawals also have a significant impact on water resources. In several countries to avoid further deterioration of streams and riparian
ecosystems, policies have been implemented to preserve a fraction of streamflow which allow ecosystem functioning. However, these were implemented due to several issues taking place. For example, the shrinkage of big reservoirs (Micklin and Aladin, 2008) or rivers discharge dropping to zero (Chen et al., 2003), or the intrusion of saline water (El-Bihery and Lachmar, 1994) are all consequences of mismanagement of surface water, which has affected different regions around the world.

The sustainability of groundwater resources has also been studied. For example Gleeson et al. (2012), based on the development of a groundwater footprint, showed how a significant fraction of world population (1.7 billion people) is under threat due to unsustainable exploitation rates of aquifers. Dalin et al. (2017), by addressing these groundwater depletion rates, links this trend to irrigation and the international food trade chain.

Therefore, water exploitation must take into account climate scenarios and should be carried out considering renewal rates and ecosystem functioning to avoid further damage to the environment, and reduce the stress and risk associated with running out of resources. Alternative management practices are also required to increase water availability in order to sustain a growing population.

However, since climate change is posing an external pressure on water resources and the ability of human population to adjust to such changes, its consequences on water resources should also be discussed.

1.2. Climate change impact on water resources

The effects of climate change on water resources have been widely described in several studies (Arnell, 1999; Vörösmarty et al., 2000; Gosling and Arnell, 2016). One of the most common conclusions is the spatially variable nature of change (Scherler et al., 2011; Collins et al., 2013). For instance, while temperature seems to be mostly increasing globally, rainfall patterns are variable (Collins et al., 2013). However, some studies have pointed out that in terms of rainfall, even if these do not change significantly in terms of their total amount, the frequency of rainfall events should diminish, and their magnitude is expected to increase (Chou et al., 2012).

These temporal rainfall pattern changes have several implications. For example, an increase in the magnitude of rainfall events may lead to increased runoff due to rainfall intensities exceeding soil infiltration rates (Siteur et al., 2014). If such a scenario takes place, then runoff magnitudes could lead to exceeding infrastructure
design capacities due to changes in intensity-duration-frequency curves, which could imply higher impacts of floods (Milly et al., 2002; Hirabayashi et al., 2013; Arnell and Lloyd-Hughes, 2014; Misra, 2014), yet these could also cause increased streamflow and groundwater interaction, resulting in higher recharge from streams. As pointed out, the relationships between the different water budget components are convoluted. Eekhout et al. (2018) for instance explains that higher precipitation intensities may cause lower soil infiltration and greater water reservoir inflows, i.e. result in a redistribution between green and blue water. However, if no new storage capacities are implemented, this can lead to an increase in water loss by runoff, reducing water security.

Additionally, if the predicted rainfall change pattern occurs, it might also lead to a reduction in available water if storage works are not improved to match the new runoff volumes. But it also implies a decrease in the volumes of rainfall infiltrated, which might cause a decrease in soil moisture, reducing the green water component and recharge from rainfall (Holman, 2005; Dunkerley, 2011). Additionally, increasing global temperatures are accompanied by a melting of ice masses, which again translates into a loss of available water (Meehl et al., 2007; Scanlon et al., 2016).

Reference evapotranspiration has shown a variable trend in different locations (Shenbin et al., 2006; Johnson and Sharma, 2010; Wang et al., 2011). Even though increasing global temperatures under climate change could lead to a higher evaporation gradient, other variables, such as radiation or humidity, may offset this increase in some areas (Wild et al., 2004; Xu et al., 2006), which leads to non-significant changes in reference evapotranspiration. However, in some regions, the increase in temperatures does cause higher evapotranspiration. For instance, most meteorological stations in Australia indicate a significant positive trend (Johnson and Sharma, 2010).

The consequences of such trends are variable. For example, higher reference evapotranspiration means higher irrigation needs for crop production (Allen et al., 2004). They also imply higher open water evaporation losses and therefore, a reduction in the available water. These can lead, if not compensated by a decrease in water withdrawals, to a reduction in surface water that can also hamper groundwater recharge. Thus, different components of the water budget have a feedback on other components, affecting the hydrologic cycle.

In addition, the melting of polar glaciers is causing an increase in sea water levels. This and a reduction in groundwater recharge, added to excessive water extraction in coastal regions may lead to higher saline water intrusion in some areas (Luoma et al., 2013; Rasmussen et al., 2013; Green and MacQuarrie, 2014).
Considering the aforementioned factors, several studies project a dramatic future increase in population exposed to water scarcity as well as hunger, due to a shortage in crop yields, all driven by climatic change (Misra, 2014; Gampe et al., 2016; Gosling and Arnell, 2016; Mekonnen and Hoekstra, 2016).

One of the expected consequences of these changes involves entering into a cycle of land degradation and poverty, or the intensification of land degradation and poverty, which can ultimately lead to the migration of entire villages (Hummel, 2015). These changes also are causing the redistribution of crops in order to match climate requirements (Ye et al., 2015). But not just crops are being redistributed in order to match new climate conditions, entire ecosystems and the ecological communities within are changing as an adaptation to these shifts (Pecl et al., 2017; Holsinger et al., 2019).

In the second chapter of the present thesis, both climate change trends and water usage are discussed to assess their impact on surface water resources in a big agricultural catchment located in NSW, Australia. This catchment, associated with the Namoi River and its tributaries, was subsequently used as an example to study different hydrological processes and potential management practices to tackle issues derived from water consumption and changing climate conditions.

In order to carry out these analyses, it is important to study the spatiotemporal patterns of different water budget components in order to evaluate changing conditions, which include climate and consumption rates. To achieve this, different alternatives for studying water resources may be implemented.

In this regard, the advent of remote sensing is a huge advance in resources monitoring, which can be coupled to monitoring station data to address the difficulty of studying spatiotemporally variable processes.

### 1.3. Remote sensing potential for hydrologic studies

Since data acquisition by direct measurements is expensive and limited to single points, indirect alternatives to address the spatio-temporal variability may be carried out (Ahmad et al., 2010; Maeda et al., 2011). While direct measurements lead to point estimates, these can hardly be extrapolated spatially to different situations. That is the reason why remote sensing is increasingly being used in different fields of science (Lettenmaier et al., 2015).

Even though remotely sensed data can take into account the spatial distribution of processes over the surface, they do not necessarily monitor specific hydrologic
properties. That is the reason why in some studies remote sensing is applied in conjunction with ground monitoring data (Hirpa et al., 2013; Cohen et al., 2018). While ground stations tend to capture information at high temporal resolutions, satellite imageries present variable temporal resolutions (Tourian et al., 2016).

Combining optical remote sensing data with direct measurements must be considered as an intermediate alternative to cope with the limitation of sparse ground data (Field et al., 1995; Seto and Kaufmann, 2003). Additionally, surface processes can be analysed at the temporal resolution of satellite imagery (Militino et al., 2018). In the case of surface water, it presents a particular spectral signature which is used for its detection through surface reflectance data (Ji et al., 2009). Even though mountain and cloud shadows can lead to a confusion of surface water, due to the common lower reflectance, terrain models can contribute to the discrimination (Guerschman et al., 2011).

Different satellites have different characteristics which makes them suitable for different purposes. Additionally, different satellites contain different sensors which capture surface reflectance data at different wavelengths, allowing for the detection of different materials (Cracknel, 2007). For instance, sensors in the Moderate Resolution Imaging Spectroradiometer (MODIS) satellites allow monitoring land surfaces, clouds, aerosols, ocean features, phytoplankton, water vapour, surface and cloud temperatures, ozone, cloud properties, among others, based on the different monitoring wavelengths (Ardanuy et al., 1991).

However, there are additional remote sensing data available, which can also contribute to the progress of geo and hydrological sciences. For instance, the Gravity Recovery and Climate Experiment consists of satellites that monitor the gravity field anomalies of Earth, and therefore allow for groundwater monitoring at low spatial resolution (Tapley et al., 2004). Synthetic Aperture Radar satellites use pulses of radio waves, which are echoed by ground surfaces and registered by the sensors and can monitor the ground surface without being affected by clouds. Therefore, SAR has been for example used to generate digital terrain models, flood monitoring, deforestation (Mcnairn and Shang, 2016). An example of SAR technology is the Surface Water and Ocean Topography (SWOT) mission which is projected for 2021, which is intended for altimetry data generation, especially for monitoring the water height of sea surface and terrestrial open water bodies (Bonnema and Hossain, 2019).

In this thesis, remote sensing data is coupled with ground station data to monitor surface water resources in the different chapters, which characterises the spatio-temporal dynamic of available water. In the second chapter, using remote sensing data, surface water was evaluated in the long-term at the catchment scale. In the third chapter, flood dynamics was analysed and related to ground data, while in the
fourth and fifth chapters algorithms for a volume quantification of surface water were studied, and applied at the catchment scale, respectively.

1.4. Groundwater recharge

Due to the increasing global demand of water resources and the necessity of integrated water resource management, it is clear that a better understanding of the hydrologic cycle is needed, and especially in relation to groundwater recharge (Nimmo et al., 2005, Scanlon et al., 2006). This could provide more accurate projections in terms of groundwater availability and aquifer sustainability (Healy and Cook, 2002). However, recharge is one of the most difficult water balance variables to estimate because of the necessity to measure it below the ground surface and its high spatio-temporal variability (Glendenning et al., 2012).

There are different factors affecting groundwater recharge, some of them subject to major changes because of human activity. One of the most obvious is climate, which ultimately defines the quantity of water in a catchment, its spatio-temporal distribution, and the evaporation potential. Thus, as earlier discussed, expected changes in rainfall frequency and intensity should also impact the amount of recharge (Eckhardt and Ulbrich, 2003). For example, whilst some authors (Scanlon et al., 2006; Zomlot et al., 2015) have stated a direct correlation between precipitation and recharge, Wang et al (2015) and Holman (2005) found that the intensity of rainfall also affected this relationship.

Other factors that should also be considered when studying recharge are vegetation, land use, topography, geology and soils. Both, vegetation and land use affect the recharge by differences in actual evapotranspiration, rainfall interception, presence and depth of roots, and the roughness of the ground (Scanlon et al., 2006, Owuor et al., 2016). Topography affects the infiltration opportunity and the amount and size of sediments overlying the bedrock (Appels et al., 2015; Ehlers et al., 2016). Soils affect the infiltration rate which can vary significantly depending on the soil texture and structure (Zomlot et al., 2015), while geology can lead to structural differences that can become preferential pathways to the groundwater (Heeren et al., 2010, Johnston, 1987).

In spite of the changes in the above factors caused by human activities, such as land use modifications and land clearance causing important modifications in recharge rates, the pumping of groundwater for irrigation, human consumption, and the development of other economic activities has significantly mined groundwater resources (Scanlon et al., 2006).
In semiarid and arid regions floods play a significant role in groundwater recharge. During intense precipitation the temporal concentration of rainfall events can even exceed the infiltration rate, causing runoff. It can also produce shorter residency times for water susceptible to evapotranspiration in soils (Nimmo et al., 2005), which will be determined by the magnitude of the flood as well (Morin et al., 2009).

However, there is a lack of understanding of flooding processes and their influence on recharge, especially when quantification is required (King et al., 2015). Thus, in a report carried out by Rassam et al. (2008) about groundwater models, the authors recognize that some critical processes, such as flooding recharge, are either usually simplistically modelled or not considered at all, and the uncertainty in the frequency of flooding events can lead to considerable modelled errors in areas where floods are important but episodic recharge mechanisms, such as in the case of arid and semiarid areas (Doble et al., 2014; Doble et al., 2012).

There is an overlap of processes occurring when considering flood recharge in groundwater models. On the one hand, flood events produce overbank flow of surface waters. This has the potential of causing short term localised recharge pulses in floodplain areas (Harrington et al., 2002), and can lead to ponded water that overlies clay/silt sediments (Doble et al., 2014). On the other hand, flood events increase the flow on perennial stream beds, increasing the stream losses as a source of groundwater recharge, even leading to ephemeral streambed flow in some zones and occasions (Morin et al., 2009; SWS, 2012).

Therefore, groundwater recharge that occurs during flood events must be considered with its own spatial variability, which lead to two recharge processes taking place, in the floodplains and in the streambeds, but at different rates.

When studying the streambed fluxes from floods, there are three exchange processes of different magnitude that are commonly taking place: hyporheic exchange fluxes, that correspond to fluxes occurring in the mixing region between subsurface water and surface water; river-aquifer exchange fluxes, which imply fluxes beyond the hyporheic zone; and bank storage exchange fluxes, that involve losses by bank infiltration during floods, which subsequently return to the stream when the stream discharge decrease (Chen and Chen, 2003; Packman and Bengala, 2000; Sophocleous, 2002).

Those exchange fluxes that occur as surface and groundwater interaction in streams can be divided into different groups. This is mainly based on the spatio-temporal scale of their occurrence, from seconds to years and centimetres to kilometres. Only river-aquifer exchange fluxes can be properly regarded as groundwater recharge because their fluxes take place at coarser spatial and longer temporal scales. Cranswick and Cook (2015), by analysing the magnitudes of these exchange fluxes
from data collected from 54 different studies, found that all these are positively correlated to the river discharge. Hyporheic exchange fluxes were found to be approximately one order of magnitude higher than river-aquifer exchange fluxes, which in turn are almost four times higher than bank storage exchange fluxes at the same river discharge.

Timms et al. (2001) stated that during floods, as the flow occurs through ephemeral channels, these start to lose definition on the plains, and flood water can remain stagnant for weeks, especially considering the low conductivity of the clay layers in flat topographies. However, the same study demonstrates that recharge does not penetrate to significant depth homogeneously in the floodplains. Besides, the authors suggest that groundwater mounds might dissipate partially by lateral drainage towards surface channels through cracks in clay layers, which could also explain a significant reduction of electrical conductivity of groundwater in the vicinity of surface channels. This theory could be supported by the study of Rassam et al. (2013) which, by analysing the surface-groundwater interaction, found that during periods with floods and droughts, rivers oscillated from gaining to losing systems depending on the groundwater fluxes, respectively, which might be associated either to bank storage or to hyphoreic exchange fluxes.

The development of clay fractures in soils, which can foster preferential flow can become a factor when dealing with clayey soils (Holland et al., 2006; Scanlon et al., 2002). However, in thick unsaturated clay layers that develop in lower catchment areas, cracks generated due to the swelling and shrinking behaviour could hardly reach the water table depth (Ruland et al., 1991). This would be even more improbable if abrupt lithological discontinuities between clay layers and sand and gravel sediments occur. Nevertheless, some other particular structural changes can also promote preferential flow in thick clay layers, allowing macropore flow and explaining the rapid response of groundwater to rainfall in comparison to soil matrix flow rates (Johnston, 1987).

In the case of the Namoi catchment, as the Namoi River has meandered through the history, which can be observed by clear current avulsion channels, it has led to significant changes in the composition of sediments. Thus, the presence of paleochannels in the floodplain produces heterogeneities that can be reflected by higher hydraulic conductivity of sediments, inducing a spatially non-uniform interaction between surface and groundwater (Heeren et al., 2010). Hence, it is expected that in some areas near the current main channel, either the clay layer has significantly lower thickness (Kelly et al., 2009), or it does not exist, but such areas are spatially constrained and their contribution to the overall recharge is not fully understood.
By analysing the different techniques to estimate the recharge some issues arise. Scanlon et al. (2002) divide the estimation recharge techniques in three main categories depending on the hydrologic source from which it comes, specifically techniques based on surface water studies, unsaturated zone studies and saturated zone studies.

The isolated use of some of the recharge estimation techniques can lead to an underestimation of recharge, especially when considering the use of surface water and unsaturated zone techniques separately, which involve the estimation of the recharge from streams and soils, respectively, but without integrating the processes already described during floods.

In addition, each of these techniques has their own disadvantages. For example, although unsaturated zone techniques can determine potential diffuse recharge, they can only be applied to small spatial scales, becoming a point estimate (Nimmo et al., 2005; Scanlon et al., 2002). Subsequently, the process of upscaling the obtained recharge value to larger scales can lead to significant errors considering the gaps that hydrogeological systems present at different scales (Domenico and Schwartz, 1998), and the occurrence of preferential flow zones. In contrast, surface water based techniques imply the estimation of localised recharge, which can overestimate the actual recharge because they neglect bank storage and subsequent evapotranspiration, but it does not consider the recharge occurring in the adjacent floodplains.

Other techniques imply the study of recharge in the saturated zone, often assumed as the actual recharge. The use of tracers and modelling are included in this category, which can be applied at different spatio-temporal scales. However, these can lead to underestimation of recharge if overbank flood is not considered (Bernard-Jannin et al., 2016). Considering the positive and negative aspects of the different techniques, some authors suggest a combination of these in order to provide higher accuracy to estimations (Nimmo et al., 2005; Scanlon et al., 2002).

Specific modelling of the overbank flood recharge, which can be critical in the application of artificial recharge on floodplains for ecosystem conservation and restoration (Doble et al., 2012), has been carried out through the use of MODIS images in some catchments of Australia. Using this method it was found that overbank flood recharge corresponds to a 4% of the total volume recharge and that there is an underestimation in the recharge values using this method compared to point-scale recharge modelling (Doble et al., 2014).

Given the high variability of conductive properties in sediments, an understanding of the underlying lithology is required. In order to generate lithological maps, geospatial information containing lithological descriptions can be further developed.
to generate 3D lithological models. For that reason, textual information must be firstly transformed into numerical data for further processing. This was initially carried out in Chapter 6 by training a natural language model, which translates words into a vectorial representation in a multidimensional space. This word embeddings model was evaluated intrinsically through three different tests and made available for other users.

Latter in Chapter 7, using the trained model on a dataset containing borehole descriptions, and combining it with a multilayer perceptron (MLP) neural network and different interpolation alternatives, 3D lithological models with their associated uncertainties were built up at relatively big scales, which realistically depict the hydrogeologic setting, and allow understanding of recharge processes.

Additionally, in Chapter 8 a site suitability map for managed aquifer recharge (MAR) was generated and evaluated using hydrograph data of groundwater monitoring wells, which allowed to recharge estimates in different points surrounding paleochannel sediments at the Namoi River. Managed aquifer recharge is further described in the next section.

### 1.5. Managed aquifer recharge (MAR)

Because of the degradation of water resources caused by human exploitation, policies and research are being developed to restore the environmental functioning of rivers and aquifers and to find alternatives to facing water scarcity. In addition, since water scarcity is also a matter of frequency, triggered by years of water surplus and scarcity fluctuating cyclically, these changes in water availability may be considered in order to develop integrated management plans. In the case of groundwater, some of the alternatives involve the replenishing of depleted aquifers using MAR.

MAR projects replenish depleted aquifers to restore their functionality, and additionally work as a rain water harvesting method (RWH), allowing harvesting of water from flood events (Hashemi et al., 2015; Scanlon et al., 2016). Projected future increases in flood events can therefore be an opportunity to store water through the recharge of aquifers. This would lead to additional water being available for agriculture and to carry out ecosystem functions. But in order to develop effective MAR, new investments and research in establishing infrastructure and effectiveness has to be done.

Additionally, socioeconomic factors also play a role in the feasibility of MAR projects (Arshad et al., 2014). By assessing the costs-benefits of MAR projects, the increase
in water efficiency compared to surficial storage structures must be considered. The high evaporation and percolation losses in surficial storage structures, which can reach up to 40% of the water stored each year, translates not only into a loss of water volume stored in a context of water scarcity, but also into economic losses, which increase the attractiveness of MAR projects (Craig et al., 2005). If climate change leads to higher evaporation losses associated to open water bodies, then MAR projects would be properly justified. However, few studies addressing this have been carried out.

There are different MAR techniques, some involve the direct recharge of aquifers through wells, including gravitational recharge of wells and the injection of water through pumps to recover it from wells located downstream in a system usually known as aquifer storage and recovery (ASR). These methods, generically known as well, shaft and borehole recharge, require usually a prior treatment of water (Arshad et al., 2014; Bouwer, 2002; Maliva and Missimer, 2012). Other systems involve the recharge of shallow unconfined aquifers (Pedretti et al., 2012) by induced infiltration directly from inundated areas or through a modification of surface features without the need of water treatment in most of the cases. These are usually known as spreading methods and include simple flooding, infiltration ponds and basins, ditches, furrows and drains and soil aquifer treatment, among other structures (Bouwer, 2002; Maliva and Missimer, 2012). Induced bank filtration takes advantage of the conductive properties of sediments in rivers or lakes to promote infiltration by abstraction of water in nearby wells. Other methods involve the modification of the stream network to promote infiltration. These are usually referred to as in-channel modifications, which include recharge dams, subsurface dams, sand dams and channel spreading (Ringleb et al., 2016).

Several MAR projects in the world are in operational stages. For instance, in the United States different MAR methods are used in several states. Scanlon et al. (2016) mention mainly infiltration MAR systems in the central valley of California with total water deliveries of approximately 14,000 Mm$^3$ between mid-1960s to 2013, and projects with a total delivery volume of 7,300 Mm$^3$ in Arizona from 1994 to 2013. Megdal et al. (2014) pointed out a huge range in scale of different operational projects in Arizona, from 0.6 Mm$^3$ yr$^{-1}$ in Chandler to 185 Mm$^3$ yr$^{-1}$ in the Tonopah Desert, while Scherberg et al. (2014) studied an ongoing MAR project in Eastern Oregon and Washington.

Glendenning et al. (2012) suggest that in India, MAR has been so widespread between the last two-three decades because of the dependency of population to groundwater, and its depletion over time, that there is even a “groundwater movement” or an “artificial recharge movement”. In other countries such as China (Liu et al., 2013), Iran (Hashemi et al., 2015), Finland (Niinikoski et al., 2016),
Chapter 1. General introduction

Tunisia (Ouelhazi et al., 2014), Jordan (Xanke et al., 2015), Italia (Mastrocicco et al., 2016), Paraguay (Magliano et al., 2015), there are some MAR projects in operational stages. In several other countries, MAR projects are still under analysis.

In the Australian case, Dillon et al. (2010) wrote about the increase of MAR projects, which in 2008 occupied five states and contributed to 45Mm³ yr⁻¹ for irrigation supplies, 7 Mm³ yr⁻¹ in non-drinkable urban water supplies and less than 1 Mm³ yr⁻¹ in drinkable water supplies, but with a potential to increase to more than 250 Mm³ yr⁻¹. In addition, studies of the stormwater quality of some MAR projects are also been carried out (Page et al., 2016).

Despite the increased interest in MAR projects, and the acceleration in implemented project, countries that have ongoing MAR projects only offset approximately 2.4% of groundwater withdrawals, while it is estimated that these recharge around 1% of global groundwater extractions (Dillon et al., 2019). Therefore, it is clear that more can be done to expand MAR implementations.

Due to the restrictions that the conductance of floodplain surface layers impose on recharge (Doble et al., 2012), and the feasibility of MAR projects depending on the infiltration rate of these layers (Arshad et al., 2014), it is clear that the location must be carefully studied in order to meet the highest possible infiltration rates, or when this is not possible, develop some method in order to increase the recharge rate bypassing the unsaturated zone.

One of the first steps in MAR is the selection of suitable sites for its implementation. Even though several studies have addressed this task using different approaches (Ringleb et al., 2016; Valdeverde et al., 2016; Sallwey et al., 2019), the selection for specific locations after mapping suitable sites has been rather vaguely defined. On the other hand, different validation schemas have been applied (Ghayoumian et al., 2007; Russo et al., 2014; Kazakis, 2018). However, no measurements of recharge have been considered as a validation source, most probably due to the difficulty of its estimation. These issues were therefore addressed in Chapter 8 of this study.

1.6. Structure of this thesis

A general schema of this Thesis with the associated methodologies and products is presented in Figure 1.1.
Chapter 1. General introduction

Figure 1.1. General schema depicting the different methodologies and products obtained in this thesis.

While Chapter 2 tries to address the potential of MAR by studying traditional water storage works and their high evaporative losses, it can also be grouped with chapters 3-5 since these address the use of remote sensing for studying the spatiotemporal variability of water and hydrologic processes at the catchment scale.

On the other hand, some alternatives for the study of the medium in which groundwater flow takes place was addressed in chapters 5 and 6. These present some innovative ways to cope with textual datasets to classify and map lithologies associated with geological descriptions of borehole datasets.

Chapter 8, by its part, evaluates the implementation of a Multi Criteria Decision Analysis technique, which together with a sensitivity analysis and validation alternatives were able to determine a potential area for the implementation of a MAR project in the Namoi catchment.

All these were followed by a general discussion addressing the different techniques implemented and their potential applicability given current regulations. Additionally, future work strategies for the continuation of the present study were also defined and followed by a conclusions section.
1.7. References


Chapter 1. General introduction


Chapter 1. General introduction


Chapter 1. General introduction


Chapter 1. General introduction


Chapter 1. General introduction


Chapter 1. General introduction


Chapter 1. General introduction


Chapter 1. General introduction


Chapter 1. General introduction


Ye, Q., Yang, X., Dai, S., Chen, G., Li, Y., Zhang, C., 2015. Effects of climate change on suitable rice cropping areas, cropping systems and crop water requirements in southern China. Agricultural Water Management 159, 35–44. doi:10.1016/j.agwat.2015.05.022

Chapter 2

Chapter 2

Long-term surface water trends and relationship with open water evaporation losses in the Namoi catchment, Australia

Abstract

Economic pressures on natural resources have led to a modification of natural surface water patterns. This has affected the hydrologic cycle at a global scale with multiple environmental and socioeconomic implications. This study evaluates long-term trends in surface water occurrence and related climate variables in a large agricultural catchment of Australia to understand trends in evaporative water losses from open water bodies. Surface water detection was based on a supervised classification on Landsat images between 1988 and 2018 to obtain inundation and surface water frequencies across the catchment, which were compared with other surface water products. Climate trends were based on monthly SILO gridded datasets. Open water evaporation was estimated using the FAO56 methodology and compared with the Penman-Monteith-Leuning Evapotranspiration V2 (PML_V2) product. Evaporation trends were analysed using the Mann Kendall (MK) test and the Sen’s slope (SS). Generally, open water frequencies showed significant negative trends, though these varied spatially. The number of dams, on the other hand, had an increasing trend. Temperatures are increasing in the catchment, while rainfall and relative humidity are decreasing, resulting in an overall positive trend for reference evapotranspiration (ETr) across 90% of the catchment. Even though ETr and evaporation per unit area of water present positive trends, lumped open water evaporation showed a negative trend, possibly associated with an average decrease in surface water frequencies. All these imply a loss of blue and green water in the catchment, and provide evidence of an overall intensification of the hydrologic cycle as predicted under climate change.
Statement of Contribution of Co-Authors

This chapter has been written as a journal article. The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
5. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for this publication is:


<table>
<thead>
<tr>
<th>Contributors</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ignacio Fuentes</td>
<td>Original idea; Coding; Data analysis; Writing</td>
</tr>
<tr>
<td>Jose Padarian</td>
<td>Edition; Suggestions</td>
</tr>
<tr>
<td>Floris van Ogtrop</td>
<td>Edition; Suggestions</td>
</tr>
<tr>
<td>R. Willem Vervoort</td>
<td>Supervision; Edition; Suggestions</td>
</tr>
</tbody>
</table>
Chapter 2. Long-term surface water trends

2.1. Introduction

The increasing anthropogenic pressure on natural resources has been widely reported (Castello and Macedo, 2015; Schandl et al., 2016; Aukema et al., 2017; Visbeck, 2018). This trend also includes water resources, which are impacted by overexploitation, affecting associated ecosystems (Castello and Macedo, 2015; Dalin et al., 2017; Russo and Lall, 2017).

Surface water is a component of the hydrological cycle, which allows both ecosystem functioning and socio-economic activities (Davis et al., 2015). However, pressures on surface waters, such as land use change and economic activities, lead to the modification of natural surface water patterns (Gerbens-Leenes et al., 2010; Castello and Macedo, 2015; Jin et al., 2015; Scanlon et al., 2017). Agriculture (i.e., food production) represents approximately 70% of blue water use (water in rivers, reservoirs and aquifers) and around 90% of consumptive water use, while domestic water consumption accounts for only 10% of water use (Hoekstra and Chapagain, 2007; Davis et al., 2015; Scanlon et al., 2017). Importantly, most water usage relies on surface water sources (Connor et al., 2009). According to FAO (2016), world surface water represents over 70% of total water withdrawals for consumptive use. Additionally, Gerten et al. (2013) estimated global blue water consumption to be on average 1,831 km3 y-1, which corresponds to 11.2% and 65.4% of accessible blue water and the freshwater planetary boundary (which if exceeded, would endanger Earth systems), respectively.

Meteorological variability results in a fluctuation of available blue water, affecting agricultural water supplies and food security. Dry periods reduce available water for irrigation, while wet years cause a water surplus, which is transferred into runoff if no water harvesting alternatives are promoted (Hashemi et al., 2015; Scanlon et al., 2016). However, this also affects the availability of soil moisture (green water), and consequently, productive evapotranspiration. In most cases, this temporal variability is addressed individually by farmers through building on-farm dams and ring tanks, while at the regional scale reservoir building is used for management (Cosgrove and Loucks, 2015; Pittock, 2016). Both alter the natural hydrological pattern and impact environmental assets (Castello and Macedo, 2015; Fowler et al., 2015).

However, other human activities can further strengthen climatic variability effects on surface water pattern modifications. For example, land use changes might lead to an intensification of the hydrological cycle, which modifies the spatial distribution and temporal dynamics of water resources (Jin et al., 2015; Davis et al., 2015). Energy needs also change the natural spatiotemporal distribution of surface water.
through the dams for hydropower generation, modifying the natural hydrological cycle on a large scale (Pringle et al., 2000). These activities have affected the sustainability of natural resources, and have also resulted in large changes in the landscape water balance (Castello and Macedo, 2015).

Several alternatives have been proposed for an integrated management of hydrologic resources to address these issues. Arguments for a more integrated approach focus on improving the sustainability of water resources and improving the environmental functioning of streams and wetlands, but also a more efficient water use, which implies reduced water losses (Poff et al., 2015). Among several alternatives to improve water storage, managed aquifer recharge (MAR) has gained attention, mainly because it involves replenishing of groundwater resources by recharging aquifers in years of water surplus, which can then be used in dry periods (Rawluk et al., 2013; Dillon et al., 2018). Additionally, MAR might reduce evaporation losses (Dillon et al., 2009; Dillon and Arshad, 2016). However, few studies have analysed long-term changes in surface water evaporation losses, particularly at regional scales (Zhao and Gao, 2019). Most studies have focused on reference and potential evapotranspiration changes, with contradictory results, depending on the study locations (Shenbin et al., 2006; Kingston et al., 2009; Wang et al., 2011; Dinpashoh et al., 2018). Since open water evaporation constitutes a response to both evaporation and open water extent, dynamic changes in these might impact regional evaporation estimates. In a warming global climate with spatially variable changes in frequency and intensity of rainfall (Westra et al. 2012; Ishak et al. 2013), such changes need to be understood to identify optimal ways of water storage.

Water reservoirs can involve large volumetric evaporation losses. For instance, Zhao and Gao (2019) reported annual evaporation losses from big reservoirs in the United States to be equivalent to 93% of the annual public water supply of the country. Craig et al. (2005) reported annual evaporation losses up to 40% of farm dam storages in Queensland and New South Wales in Australia. However, most studies refer to short-term assessments or are based on particular water reservoirs (Gibson, 2002; Craig, 2006; Liu et al., 2009). In contrast, long term changes on surface water pattern at regional scales might lead to a change in the dynamics of evaporative losses. This aspect has rarely been studied but would increase our understanding of regional water balances. Moreover, for MAR to be an acceptable alternative for water storage, we need to understand the variability in evaporative losses and how this relates to open water storage and flood occurrence.

Development of remote sensing in the last decades has improved monitoring of surface water (Palmer et al., 2015; Donchyts et al., 2016). Since water has a distinctive spectral signature, it can be detected from space using satellites that capture the surface reflectance at different wavelengths (Ji et al., 2009).
Different alternatives for water detection from surface reflectance data are possible. Surface reflectance images can be integrated with supervised or unsupervised classification algorithms to accurately detect surface water (Fuentes et al., 2019a). Even though this methodology was better than others, such as the use of water indexes alone (i.e. Munasinghe et al., 2018), it is difficult to extrapolate accurately outside the reference study areas. Alternatively, several spectral indices have been proposed for water detection in a more automated manner (Gao, 1996; Xu, 2006; Feyisa et al., 2014), but these require definition of empirical thresholds.

Even though a (freely available) synthetic aperture radar dataset at high spatial resolution (10 m) has been released (Sentinel 1, Copernicus), allowing observation of the land surface through clouds, it has only been in operation for a short time (from October, 2014), which limits long-term monitoring studies (Zhou et al., 2017). On the other hand, surface reflectance products derived from the Landsat satellites have accumulated for more than 30 years of monitoring (Donchyts et al., 2016). This dataset now offers a valuable dataset to study long-term changes in surface processes and at relatively high spatial resolution (30 m).

This study aims to evaluate the long-term trend in surface water occurrence and the associated evolution of evaporative water losses in a large agricultural catchment in Australia using more than 30 years of Landsat imagery and gridded climate data. The study area, datasets used and pre-processing steps are detailed first, followed by a description of the water detection method and the categorization of surface water bodies. Subsequently, the open water evaporation methodology is described, which is followed by the results and a discussion regarding the observed open water trends and their relation to evaporative losses.

### 2.2. Materials and methods

#### 2.2.1. Study area

The study was based on the Namoi basin, located in New South Wales, Australia (Figure 2.1). This basin was selected for this study because agriculture is strongly developed and due to its large surface area (42,000 km$^2$). Cotton is the dominant irrigated crop in the catchment, covering over 60% of the irrigated area in the basin (Green et al., 2011). Cotton is highly dependent on the availability of water resources, which has been supported by the building of three dams named Keepit, Splitrock and Chaffey. However, these dams were also designed for flood mitigation (Thurtell and Wettin, 2012).
The Namoi catchment is characterised by two main Köppen-Geiger climate types: Humid subtropical to the east and hot steppe to the west (Peel et al., 2007). Mean annual rainfall varies between 449 mm to 1135 mm, decreasing to the west. Most rainfall occurs in summer and the coefficient of variation ranges from 14.7% in the upper Namoi (high rainfall eastern areas that feed water reservoirs) to 38.6% in lower Namoi areas (low rainfall areas to the west), where most irrigated agriculture takes place. Annual pan evaporation ranges from less than 1000 mm to around 2200 mm and increasing in a westerly direction (Green et al., 2011).

The main hydrologic feature in the catchment is the Namoi river and its tributaries, which have deposited the alluvial sediments on which most agriculture takes place. The Namoi river flows to the west from the higher elevations in the east of the catchment (Welsh et al., 2014). Streamflow variability in the catchment changes depending on the river reach, with a coefficient of variation up to one order of magnitude higher than the mean streamflow.

Figure 2.1. Study area and climatic characteristics.
2.2.2. Data and preprocessing

Images from the Tier 1 surface reflectance collections of Landsat 5, 7 and 8 were used in this study. The catchment is included in 8 tiles of the Landsat data set (Figure 2.2). The data from 1986-08-19 to 2019-06-22 used in this study correspond to a total of 6,575 Landsat images (Table 2.1). The bands between the different collections (Landsat 5/7 and Landsat 8) were homogenised and clouds and shadows were removed from all images through the pixel_qa band contained in the image collection. Additionally, the normalised difference vegetation index (NDVI) and the modified normalised difference water index (mNDWI; Xu, 2006) were calculated and added as bands of the Landsat images based on:

\[
NDVI = \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + \rho_{\text{Red}}} \tag{2.1}
\]

\[
mNDWI = \frac{\rho_{\text{Green}} - \rho_{\text{MIR}}}{\rho_{\text{Green}} + \rho_{\text{MIR}}} \tag{2.2}
\]

where \( \rho_{\text{NIR}} \), \( \rho_{\text{Red}} \), \( \rho_{\text{Green}} \) and \( \rho_{\text{MIR}} \) are the reflectance for Landsat images in the wavelength range of approximately 0.7-0.9, 0.6-0.7, 0.5-0.6 and 1.5-1.7 \( \mu m \), respectively.

Figure 2.2. Landsat images covering the study catchment associated to the tiles of the Landsat satellite constellation.
Table 2. 1. Landsat tiles and images included in the study catchment.

<table>
<thead>
<tr>
<th>Path images</th>
<th>Row images</th>
<th>Number of images</th>
</tr>
</thead>
<tbody>
<tr>
<td>089</td>
<td>081</td>
<td>814</td>
</tr>
<tr>
<td>090</td>
<td>081</td>
<td>866</td>
</tr>
<tr>
<td>091</td>
<td>081</td>
<td>880</td>
</tr>
<tr>
<td>092</td>
<td>081</td>
<td>878</td>
</tr>
<tr>
<td>093</td>
<td>081</td>
<td>574</td>
</tr>
<tr>
<td>089</td>
<td>082</td>
<td>837</td>
</tr>
<tr>
<td>090</td>
<td>082</td>
<td>868</td>
</tr>
<tr>
<td>091</td>
<td>082</td>
<td>858</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>6575</strong></td>
</tr>
</tbody>
</table>

The collection was subsequently filtered to images between 1988-01-01 and 201901-01, which results in 31 years of monitoring, since images before 1988 have a lower frequency. This reduced the collection to a total of 6,351 Landsat images.

Gridded datasets for daily minimum temperature ($T_{min}$), maximum temperature ($T_{max}$), mean temperature ($T_{mean}$), mean relative humidity ($RH_{mean}$), rainfall and short crop reference evapotranspiration ($ETr$) were downloaded from the SILO webpage (https://www.longpaddock.qld.gov.au/silo/) for the study period. This daily data has a spatial resolution of 5 km and is the result of an interpolation (Jeffrey et al. 2001) of observed weather variables. The downloaded datasets were clipped to the extent of the study catchment, monthly averaged or summed, and subsequently analysed.

Additionally, the water body evaporation band from the coupled evapotranspiration and gross primary product (PML_V2) was used as a reference to evaluate the estimated open water evaporation from the SILO collection. In this dataset, Penman evaporation was considered as the actual evaporation from the water bodies and it was derived from the GLDAS collection (Rodell et al., 2004) and MODIS imagery (Zhang et al., 2019).

The smoothed Australian digital elevation model (DEM-S) was used in the water detection (Gallant et al., 2011), and discharge data from a gauging station located in the Namoi river at Goangra (in the western downstream reach of the catchment) was also used. It was downloaded from the Water NSW webpage (https://realtimedata.waternsw.com.au/).

In order to evaluate the results of the surface water classification two surface water datasets were used. These correspond to the monthly water history of the Joint
Chapter 2. Long-term surface water trends

Research Centre (JRC; Pekel et al., 2016), and the Water Observations from Space map (WofS; https://www.ga.gov.au/scientific-topics/communitysafety/flood/wofs).

Finally, major reservoirs and the main path of the Namoi River and its tributaries were obtained from the Australian Hydrologic Geofabric dataset (Bureau of Meteorology, 2012).

2.2.3. Surface water detection

A random forest model (RF) was trained with a mosaic based on eight landsat images covering the study area (Table 2.2). The images selected for training the RF classifier were based on images that coincide with flood events in the catchment, two images with high solar zenith angles within the collection to differentiate between the spectral signature of water and mountain shadows, and a few randomly selected images. The classification used all surface reflectance bands and the NDVI and mNDWI appended bands. However, since the study area contains flat and mountainous landscapes, the slope from the DEM-S was also added as a band to the Landsat images to serve as an additional input to the classifier. The hill shadow of the images was also estimated through the DEM-S together with the zenithal and azimuthal angles of the images, and this was appended as an additional band to the images. In the hill shadow mask of the images, polygons of the major reservoirs (Keepit, Chaffey and Split Rock dams, and Goran Lake) were buffered by 300 m, and assumed to be free of shadows.

Table 2.2. Images and characteristics used as variables in the random forest classifier.

<table>
<thead>
<tr>
<th>Image ID</th>
<th>Satellite</th>
<th>Date</th>
<th>Tile</th>
<th>Zenith angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC08_092081_20170428</td>
<td>Landsat 8</td>
<td>2017-04-28</td>
<td>092081</td>
<td>53.68</td>
</tr>
<tr>
<td>LT05_091081_20101213</td>
<td>Landsat 5</td>
<td>2010-12-13</td>
<td>091081</td>
<td>30.08</td>
</tr>
<tr>
<td>LE07_089081_20001211</td>
<td>Landsat 7</td>
<td>2000-12-11</td>
<td>089081</td>
<td>29.85</td>
</tr>
<tr>
<td>LE07_092081_20001216</td>
<td>Landsat 7</td>
<td>2000-12-16</td>
<td>092081</td>
<td>30.28</td>
</tr>
<tr>
<td>LE07_090081_20001218</td>
<td>Landsat 7</td>
<td>2000-12-18</td>
<td>090081</td>
<td>30.45</td>
</tr>
<tr>
<td>LT05_091082_19950627</td>
<td>Landsat 5</td>
<td>1995-06-27</td>
<td>091082</td>
<td>70.66</td>
</tr>
<tr>
<td>LT05_090080_19950620</td>
<td>Landsat 5</td>
<td>1995-06-27</td>
<td>090080</td>
<td>68.09</td>
</tr>
<tr>
<td>LC08_090082_20140811</td>
<td>Landsat 8</td>
<td>2014-08-11</td>
<td>090082</td>
<td>56.64</td>
</tr>
</tbody>
</table>
A set of 105 polygons for surface water and 245 polygons of dry land end member pixels were manually delineated from the Landsat training mosaic (Figure 2.3). Surface reflectance pixels within those polygons were sampled, and randomly split into a training subset, containing 90% of the original pixels within polygons (41,078 pixels), and a validation subset, containing the remaining 10% (4,521 pixels).

**Figure 2.3.** Surface water classification examples. The upper row of images correspond to false colour landsat images (NIR-Red-Green) of Keepit (left) and Splitrock (right) dams, which have been pre-processed. Dry land (yellow) and surface water (turquoise) polygons were manually drawn to feed the random forest classifier. The lower images correspond to surface water classified rasters derived from the upper images.

The RF classifier was trained using the training subset with one Rifle decision tree per class, a size of terminal nodes set to 1, a fraction of input to bag per tree of 0.5 and the square root of the number of variables per split.

After the classification was done, the estimated hill shadow band in the images was used to mask the classified images. This takes into account that, apart from some
of the major reservoirs which were removed from the hill shadow mask, no other water bodies are present in areas with steep terrain in the catchment.

The evaluation of the classification was based on a confusion matrix, which gives the overall accuracy and the kappa value for the classifier and for the validation subset, and also the producer and consumer accuracies. The performance evaluation was based on the training mosaic image, using the validation subset polygons, and four other randomly selected landsat images, in which 400 dry points and 100 water points were delineated. Additionally, an annual recurrence map was obtained through the sum of maximum annual inundated areas in the catchment and a perpixel probability of water occurrence map were also used as validation sources for the implemented classifier.

The JRC product was used to compare the results of the RF classified dataset. In addition, 10,000 random points were sampled from the WofS product, which only provides historical probabilities of water occurrence, to compare with the probabilities of water occurrence derived from the RF classification.

### 2.2.4. Surface water classes and trend analysis

The monthly and annual frequency of the occurrence of surface water in the classified images was analysed. This was assessed for each pixel in both the classified collection and the JRC monthly water history dataset by summing the times that the pixel appears as inundated within a year or month and dividing it by the number of times that it is within the collection in the year.

Monthly surface water frequency images were obtained for the evaluation of evaporation, since this is strongly seasonal. Gaps in the monthly surface water data were filled by linearly interpolating the surface water frequency images, assuming that surface water variation is linear.

Three classes of inundated pixels were defined for the annual water frequencies 1) permanent water bodies, with an annual frequency greater or equal than 0.75, which imply that these surfaces are inundated most of the year, 2) seasonal water bodies with a frequency smaller than 0.75 and greater or equal than 0.25, which imply that these surfaces are inundated between 1 and 3 seasons per year and 3) episodic water bodies which have an annual frequency lower than 0.25, meaning inundation during less than one season per year. Additionally, the annual average area of surface water was estimated as the sum of pixel areas multiplied by the water frequency as shown in eq. 2.3 (Xia et al., 2019):
Chapter 2. Long-term surface water trends

\[ \bar{A} = \sum_{i,j}^{N} \frac{n_{i,j}}{m_{i,j}} \]  
(2.3)

where \( \bar{A} \) is the annual lumped average surface water occurrence, \( i,j \) corresponds to the row and column of each pixel in the image, \( N \) is the total number of pixels in the image, \( a_{i,j} \) is the pixel area, \( n_{i,j} \) are the times that water occurs in each pixel during the year, and \( m_{i,j} \) are the times that the pixel occurs within the year.

On-farm dams were annually evaluated assuming that these corresponded to surface water areas with a frequency higher than 0.33 and surface inundated areas between 5,000 and 500,000 m². The resulting masked rasters were subsequently vectorised, filtering out polygons that intersected the main river and reservoirs. From those, polygons were counted annually and their total area was estimated.

Trend analysis of monthly data was carried out after a seasonal decomposition analysis to remove the seasonality in the data, while annually lumped data was directly evaluated. Trends were obtained through the Mann Kendall (MK) test and the Sen's slope (SS). These allowed quantification of the magnitude, direction and the significance of trends.

### 2.2.5. Evaporation from open water

The daily evaporation of open water bodies was calculated from the SILO gridded ET data, which uses the FAO Penman-Monteith equation (FAO56). This corresponds to the short crop \( E_{T_{r}} \) of the ASCE standardised reference evapotranspiration equation (Eq. 2.4; Walter et al., 2000; Allen et al., 2004):

\[ E_{T_{r}} = \frac{0.408\Delta(R_{n}-G) + \gamma \frac{e_{n}}{T + 273}\Delta(t_{s}-e_{a})}{\Delta + \gamma(1+C_{d}u_{2})} \]  
(2.4)

where \( E_{T_{r}} \) is the daily standardised \( E_{T_{r}} \), \( \Delta \) is the saturation vapour pressure-temperature slope, \( R_{n} \) is the net radiation, \( G \) the soil heat flux, \( \gamma \) is the psychrometric constant, \( T \) is the mean daily temperature, \( u_{2} \) the mean daily wind speed at 2 m, \( e_{s} \) the saturation vapour pressure, \( e_{a} \) the actual pressure vapour, and \( C_{n} \) and \( C_{d} \) are constant values associated to the short crops, equivalent to 900 and 0.34, respectively.

\( E_{T_{r}} \) was monthly and annually summed in order to get lumped \( E_{T_{r}} \). Evaporation of water surfaces can be calculated by scaling \( E_{T_{r}} \) with a constant crop factor (Allen et al., 1998; Craig, 2006; Jensen, 2010) as shown in eq. 2.5:

\[ E_{a} = E_{T_{r}} K_{c} \]  
(2.5)
where \( E_a \) is actual evaporation from open water bodies and \( K_c \) is the crop coefficient, which for open water bodies ranges from 1.05 to 1.25, depending on the water depth, being 1.25 for deep open water bodies (Allen et al., 2004).

Additionally, since no flux tower measurements are available (Glenn et al., 2011) in the study catchment, the PML_V2 dataset was used as a reference for validation of the estimated evaporation values. The water evaporation band from this data set was sampled at 4 different locations in known surface water bodies (Keepit, Splitrock and Chaffey dams, and Goran lake). This dataset has been recognised as one of the “state of the art” sources for actual evapotranspiration and evaporation values, and has been compared favourably with several eddy covariance flux towers (Zhang et al., 2019). Yet, the PML_V2 product was mainly intended for evapotranspiration monitoring from land surfaces, and was not validated with open water evaporation measurements. This dataset was used as a reference for validation and not directly for lumped evaporation since the time interval of the images covers 2002 to 2017, which would lead to a reduction in the analysed time series. In addition, its resolution at 500 m for evaporation of water bodies neglects several of the farm dams in the catchment which have lower spatial dimensions. These farm dams are one of the agricultural characteristics in the catchment for which we want to assess the evaporation.

For the validation, estimated evaporation values from the SILO dataset were averaged to the time scale of the PML_V2 dataset, which is based on 8-day composite images derived from MODIS.

By using the annual monthly surface water rasters and the total monthly evaporation rasters, monthly water losses by evaporation were calculated and lumped to the catchment scale, yet the spatial analysis of evaporation trends was carried out using the series of total annual evaporation.

### 2.3. Results

#### 2.3.1. Water detection

The performance of the water detection method is given in Table 2.3. Overall, the detection can be classified as very good for the different evaluated dates (kappa > 0.8 and overall accuracies > 0.9). The lowest performance can be observed using the Landsat 7 collection, followed by Landsat 8. In this case, these are associated with lower producer’s accuracy, since some water bodies are not being classified as water. However, these omission errors are only a very small fraction of the pixels actually belonging to the different classes, which explains the high accuracy and kappa values. Xia et al. (2019) also found higher errors using Landsat 7 to evaluate
surface water in China. Other issues, such as the presence of mixed pixels at the perimeter of water bodies, lead to confusion in both the user (reference data) and the algorithm. However, this is more related to the resolution of the data than a problem of the algorithm, which cannot be analysed with the current performance analysis.

Table 2.3. Surface water detection performance.

<table>
<thead>
<tr>
<th>Images</th>
<th>Kappa</th>
<th>Overall</th>
<th>Producer</th>
<th>Consumer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mosaic training</td>
<td>0.995</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>Mosaic validation</td>
<td>0.995</td>
<td>0.998</td>
<td>0.998</td>
<td>0.996</td>
</tr>
<tr>
<td>LE07/C01/T1_SR/LE07_090081_20180408</td>
<td>0.850</td>
<td>0.956</td>
<td>0.885</td>
<td>0.973</td>
</tr>
<tr>
<td>LT05/C01/T1_SR/LT05_093081_20110418</td>
<td>0.974</td>
<td>0.992</td>
<td>0.980</td>
<td>0.995</td>
</tr>
<tr>
<td>LC08/C01/T1_SR/LC08_092081_20140420</td>
<td>0.955</td>
<td>0.986</td>
<td>0.965</td>
<td>0.991</td>
</tr>
<tr>
<td>LC08/C01/T1_SR/LC08_090082_20140811</td>
<td>0.948</td>
<td>0.984</td>
<td>0.960</td>
<td>0.980</td>
</tr>
</tbody>
</table>

The annual water recurrence map, which integrates the entire collection of images is shown in Figure 2.4. Inundated areas from the recurrence map generated from Landsat images are similar to products generated from MODIS imagery (Fuentes et al., 2019b).
Chapter 2. Long-term surface water trends

Figure 2.4. Surface water recurrence patterns in the Namoi catchment.

In the map, major reservoirs in the catchment can be easily distinguished because of the high recurrence of surface water. The ephemeral Goran lake water body in the lower centre of the catchment can also be easily recognised, and the flood inundation pattern in the catchment, indicated by yellowish and greenish colours in the map, follows the main alluvium in the catchment and the associated floodplains. Since the topography in the west of the catchment consists mainly of flat areas, major inundation extents can be observed in those sectors.

A comparison with the 10,000 samples of the WofS product gives an $r^2$ of 0.91 and an RMSE of 1.8. There is a bias in the data towards the WofS product (not shown). However, WofS includes data until 2012, while the RF classification includes data from 1988-2018, with the period 2012-2018 characterised by worsening drought conditions, which might explain this shift.
2.3.2. Water categories and trends

The classification of surface water into episodic, seasonal and permanent extents is highlighted in Figure 2.5 for both the RF classified images and the JRC dataset. All surface water fractions display some degree of temporal oscillation, reflecting dry and wet periods. The episodic extent responds quite obviously to large climatic events which result in floods. For example, flood events can be observed in 1998, 2000, 2010, and 2012, which coincide with discharge peaks recorded at the gauging station in the Namoi river at Goangra, located in the western area of the catchment (Figure 2.5c).

Figure 2.5. Components of surface water and their evolution in the study period using the RF classifier (a), obtained from the JRC dataset (b) and streamflow recorded in the Namoi river at Goangra (c).

While small differences in the permanent and seasonal water fractions between both datasets can be observed, with a higher seasonal water fraction in the JRC compared with the RF classified images, strong differences in the episodic fraction are also presented. An evaluation of the classification in flood and dry dates was carried out comparing both datasets (Appendix section).

Since the JRC datasets relates to a single image per month, and even though it is based on the surface reflectance images, it misses large surface water extents during floods possibly associated with a smoothing effect. In contrast, during dry dates it leads to slightly larger water extents than the RF classified images.

For the lumped surface water fractions, significant trends were not observed (p-value ≥ 0.05) using both datasets. However, both datasets indicated negative trends for lumped average surface water, and also for permanent and episodic water fractions. The mean of annual average surface water extents for the JRC and for the RF classified images was 103.5 km² and 102.8 km², while the coefficient of variation was 50% and 40.5%, respectively.
Average surface water extent in the catchment is autocorrelated at a 1 year time lag, however, after the first year, the autocorrelation is not significant (Figure 2.6). There appears to be a 6-7 year (non-significant) oscillating cycle in the autocorrelation of the surface water, which may be an effect of major climatic cycles, such as El Niño and La Niña.

![Figure 2.6. Autocorrelation of mean surface water extent at increasing lags.](image)

Even though lumped surface water fractions did not show significant trends, the MK and SS analysis on the water frequency images detects negative trends in some of the main water reservoirs in the catchment over the study period (Figure 2.7), yet the p-value of the trend indicates clear spatial variability. Results in the main water reservoirs show strong evidence of negative trends mostly in the perimeter, while the ephemeral Goran lake indicates significant negative trends (p-value < 0.05) over the whole area.
Using the JRC dataset, an area of 106 km$^2$, equivalent to 0.25% of the catchment area, indicates strong evidence ($p$-values $< 0.05$) of water frequency trends at an average -0.005 per year. This implies an overall loss in surface water frequency and a reduction of around 2 days of inundation per year. Within this extent, 18.5 km$^2$ present strong evidence of increasing frequency trends at an average 0.017 per year. The RF classified images indicated 156 km$^2$ with strong evidence of surface water frequency changes, with an average surface water frequency change of -0.004 per year. Within this extent, 21 km$^2$ presented positive trends of surface water frequencies at an average 0.012 per year. In Figure 2.7, blue spots in the MK and SS maps indicate positive trends, which are mainly associated with on-farm dams in agricultural areas.

The presence of on-farm dams is shown in Figure 2.8. The average number of on-farm dams in the catchment is 397 using the RF classifier, covering an area of 30.1 km$^2$. Using the JRC dataset, an annual average of 557 dams was obtained, covering 38.3 km$^2$. However, these are temporally variable in both cases, oscillating depending on the total available water in the catchment. For instance, Figure 2.8
shows that during wet years, such as 1998, 2005, 2010, 2012 and 2016, which correspond to years with a water surplus, the number and extent of on-farm dams increases, while decreasing in dry years, matching the climatic fluctuation in available water (discharge or rainfall). The coefficient of variation for the number of on-farm dams and their extent in both cases averages 25.1% and 30.6%, respectively.

**Figure 2.8.** Time series of the number of on-farm dams and associated extents.

Additionally, strong evidence ($p$-value < 0.05) of increasing the number of dams occurs in both the JRC dataset and using the RF classifier, with a mean Sen’s slope of 5.25 dam y$^{-1}$. The area covered by on-farm dams increased significantly ($p$-value<0.05) in the JRC dataset, with a Sen’s slope of 0.7 km$^2$ y$^{-1}$.

**2.3.3. Reference evapotranspiration and climatic trends**

The trend analysis of the climatic variables shows an intensification in the hydrologic cycle. On the one hand, temperatures (minimum, maximum and mean) significantly increased in the last 31 years in large extents of the catchment (Figure 2.9).
Figure 2.9. Temperature trends (upper panel) and their significance (lower panel) between 1988-2018.

Minimum temperatures show strong evidence of increasing temperatures in 20,407 km² at a mean rate of 0.0022°C month⁻¹. Maximum temperatures present strong evidence of increasing temperatures in 41,719 km² at on average 0.0036°C month⁻¹. By its part, mean temperatures show strong evidence of being increasing in 41,997 km² at a rate of on average 0.0022°C month⁻¹.

On the other hand, rainfall and relative humidity (RH) are in general decreasing. However, negative trends in rainfall cover a smaller spatial extent of the catchment. As a result, there is an overall increase in \( E_T \) (Figure 2.10).
Strong evidence ($p$-value < 0.05) of decreasing precipitation takes place in 10,000 km$^2$ (~23% of the catchment extent) at a mean rate of -0.04 mm month$^{-1}$. This translates into a total loss of 148 GL of rainfall from the beginning of the study period (1988). Relative humidity indicates strong evidence of negative trends in 38,212 km$^2$ at a rate of on average -0.012% month$^{-1}$. These lead to an overall strong evidence of increase in reference evapotranspiration in 38,491 km$^2$ at on average 0.023 mm month$^{-1}$.

To evaluate the relationship between annual average surface water and annual climate variables a cross correlation was carried out. For rainfall and $ET_r$, the cross correlation with average surface water is significant for the first lags. Here rainfall is directly related, while $ET_r$ is inversely related to average annual surface water (Figure 2.11).
Figure 2. 9. Cross correlations between rainfall and surface water extent (blue), and between \( ETr \), and surface water extent (red).

2.3.4. Open water evaporation

The open water evaporation estimated through the SILO short crop ETr (FAO56) using an open water coefficient of 1.25 is well correlated with the PML-V2 evaporation dataset (Figure 2.12a). This is also shown in a selected part of the time series plot of both datasets for the period 2012 - 2015 for the Keepit reservoir (Figure 2.12b). Only a small section of the overall time series is shown to highlight the close agreement between the series.
Figure 2. 10. Scatterplot of the PML_V2 evaporation and the ASCE estimated evaporation in four different water bodies evaluated (Keepit, Splitrock and Chaffey dams, and Goran lake) (a) and part of the overall time series of open water evaporation from both datasets at the Keepit reservoir focusing on the period between 2012 and 2015 to highlight the correlation (b).

The open water evaporation estimates are correlated with the reference dataset with an $R^2$ of 0.94, a bias of -0.09 and a RMSE of 0.57 mm. As a result this was considered sufficient for subsequent analysis of lumped actual open water evaporation at the catchment scale.

### 2.3.5. Open water evaporation in the catchment

Spatial annual evaporation trends are shown in Figure 2.13. These must be interpreted as a product of the combination of surface water and ETr. Agricultural areas in the middle of the catchment have a mixed behaviour, due to the change in on-farm dams (upper panels). These present a salt and pepper effect of blue and red colours associated with negative and positive evaporation trends, respectively. However, increased evaporation trends seem to dominate the agricultural sector. Towards the east of the catchment, it can be seen how Lake Goran (lower panels) has significantly decreasing evaporation trends, which may be explained by the significant decrease in surface water frequencies, as Lake Goran is an ephemeral and mostly terminal lake.
Figure 2.11. Spatial open water evaporation trends using MK (left panel) and SS (middle panel) in the catchment, and their p-values (right panel).

Overall, around 162 km² in the catchment indicates strong evidence (\( p \)-value < 0.05) of evaporation changes, averaging a decrease of 7.09 mm \( \text{y}^{-1} \) using the JRC dataset. However, these trends are spatially variable. Within this extent, 37 km² experiences increasing evaporation trends with on average 28 mm \( \text{y}^{-1} \). Using the RF classified images, 169 km² highlights strong evidence of decreasing trends, decreasing on average by 5 mm \( \text{y}^{-1} \). However, within this area, 22 km² has increasing evaporation trends, on average by 24 mm \( \text{y}^{-1} \).

The monthly lumped evaporation for both datasets is in Figure 2.14. A strong seasonal behaviour can be observed in both signals.
Chapter 2. Long-term surface water trends

Figure 2.12. Lumped monthly evaporation in the catchment.

The mean annual evaporation with the RF classified images is 178.3 GL with a standard deviation of 90.9 GL, while using the JRC dataset the mean and standard deviation were 225.5 GL and 90 GL, respectively. The maximum annual evaporation estimated is 446.7 GL using RF and 397 GL using the JRC dataset. Both maximum annual lumped evaporation occur in 2012, when large floods took place. On the other hand, the minimum annual evaporation was 57.7 GL using the RF classified images and 90.5 GL using the JRC dataset. Both minimum lumped annual evaporation values occurred in 2007 after several years of drought. In Figure 2.14 the effect of the millennium drought (van Dijk et al., 2013) for the study area can be clearly observed. In this period, between 2001 and 2009, only a small seasonal evaporation fluctuation can be observed and evaporation values remained below average.

A seasonal decomposition of the monthly evaporation signal can be used to evaluate the evaporation trends at the catchment scale (Figure 2.15).
Figure 2.13. Seasonal decomposition of lumped evaporation using an additive model.

The deseasonalised signal of the lumped open water evaporation is decreasing at the catchment scale. The Z score of the MK test for the lumped evaporation using the RF classified images was -4.13, with a $p$-value of 3.57e-05 and a SS of 0.016 GL month$^{-1}$, while using the JRC dataset a MK Z score of -4.77 was estimated, with a $p$-value of 1.78e-06 and a SS of 0.019 GL month$^{-1}$.

Additionally, a small but significant positive trend ($p$-value < 0.05) in the evaporation-surface water ratio was detected (Figure 2.16). This can be explained by the positive trend in $ET_r$. Thus, independently of the decrease in the catchment lumped total evaporation, more water is being evaporated per unit of surface water. Therefore, under current climate conditions, more evaporative losses are occurring per unit of exposed surface water.
Evaluating the correlation between the three fractions of surface water (episodic, seasonal and permanent) suggests that at least permanent and episodic water fractions were significantly ($p$-value < 0.05) correlated (Spearman correlation) with annual lumped water evaporation. Correlation coefficients of 0.95 and 0.93 exist between permanent water and annual lumped evaporation for the RF classified images and the JRC dataset, respectively; and between the seasonal water fraction and the lumped evaporation for the RF classified images the correlation was 0.53 while for the JRC dataset it was 0.45. Since episodic water exists in the landscape for a smaller fraction of time, the contribution to open water evaporation is smaller, despite the much greater extent.

### 2.4. Discussion

Overall, this study indicates that there is a clear change in the landscape water balance over the study period (1988 - 2018). While temperature has strongly increased, rainfall and RH have decreased on approximately 23% and 92% of the catchment extent, respectively. In contrast, reference ET (ETr) has increased over a large part of the catchment. Open water evaporation is a significant part of the catchment water losses, and this is correlated with permanent and seasonal surface water bodies. Evaporation losses are spatially variable, decreasing in Lake Goran in
the upper Namoi. In contrast, they are increasing in agricultural sectors of the catchment due to the presence of on-farm dams.

All trends in this study need to be considered in relation to the issues raised by Serinaldi et al. (2018) related to interpreting the results of null hypothesis significance tests. Therefore, trend values estimated are limited to the study catchment and study period (1988-2018), and should not be used to model future conditions or be extrapolated beyond the study area. The trend tests are preliminary screening tools, further analysis using knowledge of physical processes taking place in the study area should be done to confirm that the trends make sense. Furthermore, the use of annual summaries for the trend estimations can lead to underestimation of the temporal dependencies in the data. However, as more than 5 years of data was available, the approach was considered valid (Whiters and Nadarajah, 2015).

There are several sources of uncertainty that should be considered in this study. Landsat images have an average recurrence of 16 days per satellite, which limits water detection to days when the satellite passes (Fuentes et al., 2019b). Furthermore, clouds and shadows can limit the detection of surface water, especially after anomalous weather events, such as storms. This leads to a loss of information on surface water extent monitoring (Schumann and Moller, 2015).

Additionally, the classification step may also lead to water detection uncertainties. Difficulties in the water recognition may be due to mixed pixels transfer errors or due to the classification algorithms (Fuentes et al., 2019a). As mentioned by Feyisa et al. (2014), mixed pixels can be an important source of error, especially at large scales. Furthermore, detection errors in steep areas as a result of elevation shadows, whose spectral signature resembles water, may be a potential source of errors during the classification (Feyisa et al., 2014), yet by using the hill shadow mask, we believe these errors were reduced. Even though clouds and cloud shadows were removed from landsat images using the CFMask algorithm (Foga et al., 2017), remnant clouds and cloud shadows may still be observed in some images, which affects the surface water detection, and contributes to the uncertainty in the classification (Xia et al., 2019). The different wavelength ranges in the sensors of the different Landsat satellites also introduce some noise in the classification. This might result in classification errors, which may add to the errors derived from the Scan Line Corrector fail in the Landsat 7 satellite.

Other sources of uncertainty that may be propagated into the results are associated with the reference datasets used. For instance, the SILO ETr dataset has some limitations that should be considered. Even though it is a well recognised dataset developed using the well established FAO56 formulae (Allen et al., 1998; Eq. 4), the dataset assumes a constant wind speed of 2 m s-1, which might cause errors in the
evaporation calculation. It also might propagate the uncertainties derived from the forcing data (temperature, relative humidity, and radiation). In the case of the JRC dataset (Pekel et al., 2006), it tends to miss significant surface water during floods, and therefore would lead to underestimation of evaporation in these events. Therefore, and regardless of the value of the JRC dataset for global scale use, its application at regional scales for monitoring episodic events such as floods must be carefully considered.

The evaporation layer contained in the PML_V2 dataset (Zhang et al., 2019), used to validate our evaporation estimates, implies another source of uncertainty since it was primarily developed to obtain Penman-Monteith-Leuning Evapotranspiration. Evaporation in the PML_V2 was estimated using the Penman equation, but not validated. All flux towers for evaporation and transpiration validation are located in land areas, and therefore, do not measure open water evaporation. However, open water evaporation is rather seldom measured, and therefore this hinders the validation of results. This is the main reason for using the PML_V2 evaporation in this study.

Additionally, the use of FAO56 evapotranspiration and a single open water coefficient results in an oversimplification of the evaporation estimation (Allen et al., 1998; Jensen, 2010), but is used here in order to monitor evaporation at the catchment scale and is not constrained to the main reservoirs. Thus, the open water coefficient used is only acceptable if the surface water bodies are generally deep enough. However, since the heat storage capacity of a water body is mainly influenced by its depth (Finch and Hall, 2001), different water body depths will lead to different coefficients. As mentioned, the ETr calculation also has some limitations, for example the assumption of no heat flux. As a consequence, heat storage was not considered in this approach. However, Finch and Hall (2001) explain that for annual evaporation calculations heat storage plays a small role in total evaporation, since seasonal changes tend to cancel each other. Thus, the introduced error might be small. Due to the uncertain quality of the reference data used for validation, and the reasons aforementioned, evaporation values should be regarded only as rough estimates. Clearly, there is a need for more evaporation measurements to build and validate open water evaporation models.

Climatic trends suggest an intensification of the hydrologic cycle due to warmer conditions, which translates into higher ETr in the catchment (Huntington, 2006), and a higher evaporation per unit of surface water. In contrast to other studies that show decreasing or no trends in ETr (e.g. Shenbin et al., 2006; Wang et al., 2011), this study shows spatially variable but significant positive trends covering a large catchment extent. This agrees with Johnson and Sharma (2010) study, in which pan evaporation trends were found to be increasing at most weather stations in Australia. The increase in ETr can be linked to significant positive temperature trends across
the entire catchment, which are not being offset by other weather variables. For example, in most of the catchment there is no trend in rainfall. Together, these results imply a decrease in blue and green water availability in the catchment for ecological functioning and human needs, including agricultural and domestic water usage. These results also agree with the predicted effects of climate change due to global warming (Collins et al., 2013).

As a result of temperature and ETr trends, the annual frequency of surface water occurrence was also found to be mostly decreasing. If the loss in water frequency is confirmed as a long term trend, then this can ultimately affect ecosystem functioning and led to a degradation of local economies and higher food security risks (Mclaughlin and Kinzelbach, 2015). The Namoi catchment hosts over 46,000 hectares of wetlands, where regular flood events are essential for ecosystem health (every two years; Green et al., 2011). Spatial negative trends in surface open water frequencies could signify a decrease in the health of these systems.

Several studies have reported a decline in the extent of open surface water bodies as a consequence of anthropogenic activities (Legesse and Ayenew, 2006; Micklin and Aladin, 2008; Song et al., 2012). In the current study, such activities were not directly analysed. Even though open water areas with strong evidence of changes in frequency had mostly negative trends, several spots indicate the opposite behaviour, yet their spatial extents are small. These correspond to on-farm dams for irrigation purposes, which have been built to secure water for irrigation. Additionally, in spite of climate conditions and water availability fluctuations, on-farm dams indicated positive trends. However, water usage by irrigation was not evaluated in this study. Therefore, anthropogenic effects on the overall surface water would require additional water usage data to be conclusive.

Lumped water evaporation at the catchment scale in the study period has decreased despite an increasing ETr trend. This implies that open water evaporation is dominated by the surface water occurrence and frequency. Total annual evaporative losses in the catchment average 201.9 GL using both datasets evaluated, which corresponds to 47% of the storage capacity of Keepit dam (Preece and Jones, 2002), and exceeds the annual total account usage of surface water in the catchment (Burrell et al., 2018). Since most studies address the impact of climate on evaporation, it is still unclear how these changes in open water evaporation could feedback to climate.

Two main conclusions concerning the water footprint may be drawn based on our findings. Firstly, the results reflect a decrease in the availability of blue water (open water frequency), but also in rainfall and RH. Secondly, there is an increase in evaporation losses per unit water, associated with the general increase in temperature. This is also associated with a positive trend in ETr resulting in a
decrease in storage of green water. Overall these results imply a loss of available water, which is a concern for both ecological assets and irrigation activities in the catchment. In this study, groundwater was not analysed, however Kelly et al. (2014) reported dropping levels in several monitoring boreholes in the catchment, which also suggests a loss of groundwater resources, possibly due to the reduction in surface water availability. This suggests an urgent need to review water management taking into account the increased loss of water in evaporation.

2.5. Conclusions

The change in open water evaporation was evaluated in the Namoi catchment from 1988 to the present through climate variables and surface water detection. Overall, surface water frequencies have mostly been decreasing, yet these are spatially variable in the catchment.

Climate variables clearly indicate the predicted intensification of the hydrologic cycle as a result of climate change. Temperatures have significantly increased in the catchment, while rainfall and relative humidity indicate negative trends in large areas. Together, these have led to significant increases in ETr across 90% of the catchment extent. Additionally, rainfall and ETr were significantly cross correlated to the surface water extent.

Annual lumped open water evaporation losses average more than one third of the main Keepit dam storage capacity in the catchment and exceed the annual total usage of all surface water accounts. However, these losses are spatially and temporally variable, with decreasing trends. This might be associated with a loss in surface water frequency, which partly offsets the increase in ETr. A key problem for future water management is the increased evaporation per unit surface water extent, which might be addressed by the implementation of MAR projects.
2.6. Appendix

Figure 2A. 1. Comparison of surface water in both the JRC dataset and the RF classified images evaluated at Bogabri for three different dates.

Figure 2A. 2. Comparison of surface water in both the JRC dataset and the RF classified images evaluated at Goangra for three different dates.
Chapter 2. Long-term surface water trends

Figure 2A.3. Comparison of surface water in both the JRC dataset and the RF classified images evaluated at Wee Waa for three different dates.

2.7. References


Chapter 2. Long-term surface water trends


Chapter 2. Long-term surface water trends


Chapter 2. Long-term surface water trends

CSU/ARS Evapotranspiration Workshop.


Chapter 2. Long-term surface water trends


Chapter 2. Long-term surface water trends


Chapter 2. Long-term surface water trends


Chapter 3

Spatiotemporal evaluation of inundated areas using MODIS imagery at a catchment scale

Abstract

Accurate spatiotemporal analysis of surface water is important for managing water resources. The main objectives of this study were: 1) to assess different remote sensing algorithms’ ability to accurately determine flood extent using MODIS imagery, and 2) to study the temporal dynamics of inundated areas and corresponding return periods in a large catchment in New South Wales, Australia. The different water detection alternatives used included the Open Water Likelihood (OWL) algorithm, the Brakenridge algorithm for single images (BRAK) and composite images (BRAKC), the Islam methodology (ISL), a combination of Normalized Difference Indices (NDIS), and an ensemble of these. Landsat Thematic Mapper (TM) images were used as a reference to benchmark the performance of the algorithms. Occurrence probability maps derived from inundation images and estimated flood extent frequency curves were compared against the Global Surface Water (GSW) dataset of the Joint Research Centre. The ISL methodology indicated the best performance compared with “ground truth” inundated images derived from Landsat, while the temporal performance of the BRAK algorithm indicated significant noise based on daily imagery. The remaining approaches generally matched stream discharge peaks. Maximum cross-correlation coefficients between daily discharge and inundated areas were calculated for different time lags, indicating high spatial variability across the catchment. Accumulated rainfall accounted for up to 40% and 60% of the variation in inundation extent in daily and composite images, respectively. The ensemble technique outperformed the other algorithms to describe the dynamics of flooding, followed by the NDIS and ISL algorithms. The occurrence probability images and flood frequency curves indicated a higher inundation extent than the GSW product for return periods greater than two years, which can be explained by the different temporal resolutions. The results of this study can be applied in water resources planning and flood mitigation efforts.
Statement of Contribution of Co-Authors

This chapter has been written as a journal article. The authors listed below have certified that:

6. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
7. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
8. there are no other authors of the publication according to these criteria;
9. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
10. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for this publication is:


<table>
<thead>
<tr>
<th>Contributors</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ignacio Fuentes</td>
<td>Original idea; Coding; Data analysis; Writing</td>
</tr>
<tr>
<td>Jose Padarian</td>
<td>Edition; Suggestions</td>
</tr>
<tr>
<td>Floris van Ogtrop</td>
<td>Edition; Suggestions</td>
</tr>
<tr>
<td>R. Willem Vervoort</td>
<td>Supervision; Edition; Suggestions</td>
</tr>
</tbody>
</table>
3.1. **Introduction**

Global climate change is likely to impose several constraints on the environmental suitability of human settlements. Increasing global temperatures under climate change may result in increasing water demand, which will be accompanied by a decrease in freshwater volumes because of the melting of continental ice masses (Meehl et al., 2007; Scanlon et al., 2016). However, precipitation intensity is also expected to increase, thereby leading to increasing flood magnitudes even in areas where a decrease in flood frequency is expected (Arnell and Lloyd-Hughes, 2014; Hirabayashi et al., 2013; Misra, 2014). This increase in flood risk will not only have a direct catastrophic effect on the population, but may also affect water availability. For instance, several studies have demonstrated that flood events are important for groundwater recharge in arid and semiarid areas (Benito et al., 2010; Doble et al., 2014; Eckhardt and Ulbrich, 2003; Harrington et al., 2002). This may increase pressure on agriculture and ecosystems that are dependent on these floods. Therefore, improved understanding of flood dynamics will help improve the management of water scarcity (Meehl et al., 2007; Mueller et al., 2016; Ticehurst et al., 2015a).

There are different methods to estimate flood frequency and extent. While gauging stations allow a direct estimate of flood occurrence, they provide little information about the spatial distribution of inundation (Huang et al., 2014). In contrast, hydrodynamic models retrieve spatial and temporal flood information, yet are computationally expensive for large catchments at high spatiotemporal resolutions (Ticehurst et al., 2015a). This indicates remote sensing as an interesting alternative to evaluate flood dynamics, which can also be used to validate and calibrate hydrological and hydrodynamic models (Schumann and Moller, 2015).

Given that water absorbs a significant fraction of the light in infrared wavelengths, yet less of the other wavelengths compared with other landscape features, its spectral signature can be used in remote sensing to detect water in the landscape (Ji et al., 2009). Surface reflectance products, such as those derived from MODIS and Landsat satellites, contain information in the visible and infrared wavelength ranges, and so may be used to detect water in flooded areas.

The Landsat project has been operating since the early 1970s. Landsat images have a relatively high spatial resolution (mainly 30 m), yet at a low temporal resolution (16 days on average). In contrast, the MODIS satellites, Aqua and Terra, have been operating since 2002 and 2000, respectively. Their images have a moderate spatial resolution (≥ 250 m) and a daily temporal resolution. Landsat images tend to be better for water detection because of the higher spatial resolution; however, the higher temporal resolution of MODIS images renders them more suitable for...
studying dynamic processes, such as floods (Sakai et al., 2015; Schnebele et al., 2014).

To take advantage of the positive aspects of Landsat and MODIS, it is possible to couple both products. For example, Dao and Liou (2015) and Ticehurst et al. (2015a) used Landsat as a “ground truth” to estimate the accuracy of MODIS for flood detection. Another alternative is to downscale the MODIS imagery through spatiotemporal fusion models (Heimhuber et al., 2018; Zhang et al., 2014). However, given the high computational requirements to apply these models, their use has been limited to small areas, short time spans, and low temporal resolutions.

One of the most common approaches for surface water detection is the use of water indices and thresholds (Rokni et al., 2014; Zhou et al., 2017). While the Normalized Difference Water Index (NDWI) (Gao, 1996) appears to accurately detect surface water based on the range of near-infrared (NIR) and shortwave infrared (SWIR) wavelengths, it has also been highlighted that its performance decreases when water on the surface is found in combination with other features, such as vegetation and buildings, which are often found in flood waters (Singh et al., 2015). Therefore, other indices have been developed to improve water detection under such conditions (Chandrasekar et al., 2010; Danaher and Collett, 2006; Feyisa et al., 2014). The modified NDWI (mNDWI) is one example of these (Xu, 2006).

In addition to the water indices mentioned above, other alternatives for surface water detection from reflectance images have been developed. These require either: 1) an empirical statistical model that combines the surface reflectance bands with terrain information (Guerschman et al., 2011), 2) a combination of water and vegetation indices in a series of decision trees (Islam et al., 2010), or 3) a combination of thresholds for the different surface reflectance bands (Nigro et al., 2014).

Although flood extents have been derived from MODIS imagery using different algorithms (Guerschman et al., 2011; Islam et al., 2010; Nigro et al., 2014), there is still a need to better understand the ability of these algorithms to reproduce the spatial and temporal flooding dynamics accurately. To date, no studies have compared the performance of these algorithms in both the temporal and spatial domain. Moreover, the inundated extents allow the comparison of specific return periods, which can be contrasted with gauge data. Further, in combination with rainfall estimates, satellite inundation extent can characterize the space–time variation in flood dynamics at different gauging stations.

Therefore, using MODIS imagery, this study compares the performance of different remote sensing algorithms to determine flood extents, as well as their ability to describe the space–time dynamics of flood inundation at a catchment scale. Subsequently, we compare the flood return periods by combining remote sensing
Chapter 3. Spatiotemporal evaluation of inundated areas

and stream gauge data to better understand the catchment response to floods. This in turn can aid water resources planning and flood mitigation projects.

3.2. Materials and methods

3.2.1. Study area

The case study area is the Namoi River basin, located in the state of New South Wales (NSW) in Australia. The catchment is approximately 42,000 km\(^2\) in area. The main rivers are the Mooki, Peel, Manilla, McDonald, and Namoi Rivers. The catchment has three major water storage dams—Chaffey Dam, Split Rock Dam, and Keepit Dam—and a natural ephemeral lake, the Goran Lake. Rainfall in the catchment ranges from 400 mm to 1,300 mm per year. Its spatial distribution changes with elevation, and decreases to the west. The annual precipitation at Gunnedah (in the center of the catchment) is highest in summer and lowest in autumn. Evaporation increases from east to west, ranging from approximately 1,000 mm to over 2,200 mm per year (Green et al., 2011).

Floods occur frequently in the catchment and are a natural feature of the flat floodplain area. Floods inundate approximately half of the wetlands (46,398 hectares) in the catchment, on average, every two years (Green et al., 2011). Floods also periodically inundate the town of Narrabri, causing significant economic losses (URS, 2014). Further, floods in January/February 2012, November 2011, and June 2011 affected the towns of Tamworth, Armidale-Dumaresq, Guyra, Gunnedah, Narrabri, and Walcha.

3.2.2. Data and preprocessing

Gauging station data for surface water was obtained from the Office of Water of the NSW Government (Table 3.1). In this study, 71 hydrological gauging stations were initially used, which provided daily information of the discharge rate and water level in the streams (Figure 3.1).
### Table 3.1. Sources of information used in the study.

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauge station data</td>
<td>Daily time series of stream discharges and levels from different gauges in the basin.</td>
<td>WaterNSW</td>
</tr>
<tr>
<td>Gridded rainfall</td>
<td>SILO rasters of daily precipitation in the basin.</td>
<td>SILO</td>
</tr>
<tr>
<td>DEM-H MOD09GA</td>
<td>Australian SRTM DEM-H. Daily MODIS Terra surface reflectance, 500 m.</td>
<td>Geoscience Australia Land Processes Distributed Active Archive Center, National Aeronautics and Space Administration (LP DAAC NASA)</td>
</tr>
<tr>
<td>MOD09A1</td>
<td>Eight-day composite MODIS Terra surface reflectance, 500 m.</td>
<td>LP DAAC NASA</td>
</tr>
<tr>
<td>MOD09GQ</td>
<td>Daily MODIS Terra surface reflectance, 250 m, for Bands 1 and 2.</td>
<td>LP DAAC NASA</td>
</tr>
<tr>
<td>MOD09Q1</td>
<td>Eight-day composite MODIS Terra surface reflectance, 250 m, for Bands 1 and 2.</td>
<td>LP DAAC NASA</td>
</tr>
<tr>
<td>MYD09GA</td>
<td>Daily MODIS Aqua surface reflectance, 500 m.</td>
<td>LP DAAC NASA</td>
</tr>
<tr>
<td>MYD09GQ</td>
<td>Daily MODIS Aqua surface reflectance, 250 m, for Bands 1 and 2.</td>
<td>LP DAAC NASA</td>
</tr>
<tr>
<td>MOD44W</td>
<td>Land water mask derived from MODIS and SRTM.</td>
<td>LP DAAC NASA</td>
</tr>
<tr>
<td>Landsat 5</td>
<td>United States Geological Survey (USGS) Landsat 5 surface reflectance.</td>
<td>USGS</td>
</tr>
<tr>
<td>GSW</td>
<td>GSW Mapping Layers v1.0 (1984-2015).</td>
<td>Joint Research Centre</td>
</tr>
</tbody>
</table>

Daily rainfall from 2000 to 2017 was obtained from the Scientific Information for Land Owners (SILO) gridded rainfall data at 0.05° resolution (http://www.longpaddock.qld.gov.au/silo), and summed over all the pixels in the basin, which resulted in a time series of daily lumped rainfall over the catchment. The Australian one-second Shuttle Radar Topography Mission (SRTM) Hydrologically Enforced Digital Elevation Model (DEM-H) (Geoscience Australia, 2015) was used to calculate a multiresolution index of valley bottom flatness (MrVBF) map of the catchment, which was used as an input for the Open Water Likelihood (OWL) algorithm.
This study used MODIS daily reflectance images of the Terra satellite with a 500 m spatial resolution at daily frequency and as an eight-day composite (products MOD09GA and MOD09A1, respectively). It also used the MODIS 250 m resolution, daily and eight-day composite reflectance datasets (MOD09GQ and MOD09Q1), which contained Bands 1 and 2 (red and NIR). In the case of the Aqua satellite images, the daily surface reflectance products (MYD09GA and MYD09GQ) were only used for the composite image using the Brakenridge algorithm for single images (BRAK) because of a higher presence of clouds (Ticehurst et al., 2015b) and the malfunctioning of some sensors in the range of the SWIR wavelengths (Li et al., 2014). For this period, Landsat Thematic Mapper (TM) surface reflectance images were also used.

The Global Surface Water (GSW) mapping recurrence layer v1.0 (1984–2015) (Pekel et al., 2016) from the Joint Research Centre was used as a reference to compare the occurrence probability maps obtained with MODIS, rather than the Water Observations from Space (WOfS) product (Mueller et al., 2016), because the GSW has a high spatial resolution, is accurate, has global coverage, and is available in the

Figure 3. 1. Namoi catchment location and spatial distribution of gauge stations.
Google Earth Engine platform. This product corresponds to the frequency with which water is found in a specific location from year to year, and was derived from all the Landsat scenes between 1984 and 2015.

In addition, the MOD44W product was used to mask permanent surface waters when flood extent was analyzed relative to rainfall and discharge values in order to use just those areas that are periodically inundated. All images were managed and processed through the Google Earth Engine platform (https://code.earthengine.google.com/).

For both the MODIS imageries and Landsat TM imagery, masking of clouds and cloud shadows was done using the StateQA Bitmask and the pixel_qa Bitmask bands, respectively. The StateQA Bitmask band contains information corresponding to the different classes detected for the MODIS sensors, and the pixel_qa Bitmask contains the results of applying the CFmask algorithm on Landsat images (Foga et al., 2017). In the case of the Landsat images, a focal median was applied with a window size of three pixels to reduce noise. The MODIS daily images were subsequently masked to reduce the noise at the edge of tiles and in stripes caused by the malfunctioning of the SWIR sensors. In addition, a mask based on the aerosol quantity was applied because it was found that fires affected the performance of the algorithm, detecting water when there was none (Ticehurst et al., 2014).

### 3.2.3. Water detection methods

As indicated, to determine the flood extent with MODIS images, different algorithms were compared (Table 3.2).

**Table 3.2.** Different surface water detection algorithms used in conjunction with MODIS imagery.

<table>
<thead>
<tr>
<th>Abbreviation used</th>
<th>Brief description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OWL</strong></td>
<td>The OWL algorithm uses MODIS surface reflectance and an MrVBF map.</td>
<td>Guerschman et al. (2011)</td>
</tr>
<tr>
<td><strong>BRAK</strong></td>
<td>The Brakenridge algorithm combines Bands 1, 2, and 7 of MODIS surface reflectance products.</td>
<td>Nigro et al. (2014)</td>
</tr>
<tr>
<td><strong>ISL</strong></td>
<td>The Islam algorithm (ISL) performs a decision tree based on the Enhanced Vegetation Index (EVI) and Land Surface Water Index (LSWI) obtained from MODIS surface reflectance imagery.</td>
<td>Islam et al. (2010)</td>
</tr>
</tbody>
</table>
Chapter 3. Spatiotemporal evaluation of inundated areas

**NDIS**

Normalized Difference Indices (NDIS) combines the NDWI and Normalized Difference Vegetation Index (NDVI) obtained from MODIS surface reflectance products.

Dao and Liou (2015)

The OWL algorithm was developed for Australia by Guerschman et al. (2011). It calculates the likelihood of water in a MODIS pixel using Eq. 3.1:

\[
fw = \frac{1}{1+\exp(z)}
\]  

(3.1)

where \(z\) can be formulated according to Eq. 3.2:

\[
z = \beta_0 + \sum_{i=1}^{5} \beta_i x_i
\]  

(3.2)

Here, the original parameterization was used, and \(\beta_0 = -3.4137\), \(\beta_1 = -0.000959\), \(\beta_2 = 0.0041\), \(\beta_3 = 14.1927\), \(\beta_4 = -0.4304\), \(\beta_5 = -0.09619\), \(x_1\) corresponds to the SWIR Band 6 of MODIS images, \(x_2\) corresponds to the SWIR Band 7 of surface reflectance images, \(x_3\) is the NDVI, \(x_4\) is the NDWI (Gao, 1996), and \(x_5\) corresponds to the MrVBF obtained from a STRM DEM image. Pixel values \((fw)\) range from 0 to 1, and indicate the probability of finding water on the surface.

A second algorithm is the Brakenridge algorithm (BRAK) for water detection with MODIS images (Nigro et al., 2014), which can enable access to near real-time flood maps at a global scale. This algorithm has the advantage of using the MOD09GQ and MOD09Q1 datasets, which have a higher resolution (250 m). The NIR and red bands of MOD09GQ and MOD09Q1 are combined with the SWIR band of the lower-resolution MOD09GA and MOD09A1 products, considering a series of threshold values, and can be applied to both single and composite images (BRAKC, using the Aqua products MYD09GA and MYD09GQ and a three-day window size). However, the use of single daily images to obtain flood images does not provide a standard product from the Near Real-Time Global Flood Mapping webpage (http://oas.gsfc.nasa.gov/floodmap). Surface water is detected when the following conditions are met:

\[
\frac{\text{Band}_2 + 13.5}{\text{Band}_1 + 1081} < 0.7 \quad \text{and} \quad \text{Band}_1 < 2027 \quad \text{and} \quad \text{Band}_7 < 676
\]

Another approach is the algorithm modified by Islam et al. (ISL) (2010), which combines the Land Surface Water Index (LSWI) and the Enhanced Vegetation Index (EVI) in a decision tree to obtain flooded pixels. In this case, inundated pixels are identified after removing pixels with EVI values greater than 0.3 and with a difference in values between EVI and LSWI (DVEL) greater than 0.05. Subsequently, inundated pixels are identified from the remaining pixels with EVI values lower than or equal to 0.1 (Islam et al., 2010).
These algorithms were also compared with a commonly applied technique that uses NDIS (Dao and Liou, 2015; Zhou et al., 2017) with thresholds to detect surface water. In this case, the NDWI and NDVI were combined, where NDWI greater than 0.01 and NDVI lower than 0.2 were assumed to indicate flooded areas. The NDVI threshold was empirically selected based on an improved performance of the classification compared with the “ground truth” Landsat inundation images.

In addition, since multi-model combinations can diminish prediction errors (Li and Sankarasubramanian, 2012), an ensemble of the different algorithms was created (multi-model average). In this case, for each day, the result of the different algorithms was summed and divided by the number of algorithms. Flooded pixels were then assumed for values greater than or equal to 3 (classified as inundated in at least three of the five algorithms used).

### 3.2.4. Performance assessment

Landsat and MODIS image pairs taken on the same date were selected based on a window of 10 days after and before the peak of 90th percentile discharge events measured at the Walgett gauge station (furthest downstream gauging station in the catchment). This approach constrained the image pairs to the period November to January, when most of the rainfall occurs in the catchment. Flood extent images during flood periods were generated by applying the mNDWI algorithm to the Landsat TM images. These images were used as the baseline for evaluating the flood extent.

Applying the different water detection algorithms (Table 2) allowed comparison of the flood extents obtained from MODIS with the “ground truth” Landsat flood extents for the same dates through a confusion matrix (or error matrix) (Congalton, 1991). The confusion matrix enables calculation of the overall accuracy—that is, the ratio between correctly classified pixels and the total number of pixels (Eq. 3.3):

$$P_o = \frac{C_p}{n}$$

where $P_o$ stands for the overall accuracy, $C_p$ are the correctly classified pixels, and $n$ is the total number of pixels.

The confusion matrix can be used to derive the kappa coefficient (Eq. 3.4) (Chen et al., 2014), which considers the random probability of agreement between the reference and classified image (Eq. 3.5):

$$K = \frac{P_o - P_r}{1 - P_r}$$
where $K$ is the kappa coefficient, $P_r$ is the probability of random agreement, $a_1$ and $b_1$ correspond to inundated pixels in both MODIS and Landsat images, and $a_0$ and $b_0$ are non-flooded pixels in both sources.

Commission and omission errors can also be estimated, which correspond to pixels that are classified as a specific class, yet are not actually that class, and pixels of a specific class that were not classified as such (Bastarrika et al., 2011).

Reference values for kappa coefficients are such that $kappa > 0.8$ reflects a strong agreement, $0.8 \geq kappa > 0.6$ corresponds to a substantial agreement, $0.6 \geq kappa > 0.4$ can be interpreted as a moderate agreement, and $kappa \leq 0.4$ indicate a poor agreement (Chen et al., 2014).

### 3.2.5. OWL threshold

The threshold to maximize the performance of the OWL algorithm was determined empirically by selecting the median of the maximum kappa values calculated for each image (Figure 3.2), instead of using the overall accuracy performance measure. This was undertaken because the different pixel classes between inundated and dry were unbalanced, and this greatly influenced the overall accuracy performance measure. At the scale of Landsat images, the values calculated by the algorithm in the pixels without surface water were generally more than one order of magnitude greater than the values in the inundated pixels, thereby making the selection of an accurate threshold value difficult (Ticehurst et al., 2015a). As shown in Figure 3.2, the kappa curves from the OWL algorithm generally had a maximum around a $f_w$ threshold of 0.8, with the highest median at 0.87—a value that was subsequently used in this study to identify flooded areas.
3.2.6. Flood dynamics and frequency analysis

To evaluate the spatial and temporal dynamics of flood extents, this study calculated cross-correlations between the catchment-wide derived flood extents and the daily lumped rainfall over the catchment. Here, the catchment average rainfall was used for two reasons. First, it was used because we were interested in the flood extent for the whole catchment, rather than for individual sub-catchments. Second, it was used because the rainfall totals decrease rapidly going downstream in the catchment. As a result, local rainfall events at the downstream end of the catchment are unlikely to cause major flooding. In contrast, catchment-wide events, or events in the upper catchment, can cause widespread flooding across the catchment. Given that these upstream rainfall events ultimately also reach the lower catchment, using catchment lumped rainfall was reasonable.

Following the cross-correlations, regressions were undertaken between backward cumulative daily lumped rainfalls at increasing time lags of up to 30 days, and catchment-wide flooded extents (obtained from daily and eight-day composite images). The flood dynamics were also assessed using the maximum cross-correlation values between flood extents and the daily discharge at the different gauging stations in the catchment.
Permanently and episodically inundated areas were analyzed using a frequency distribution. For each year, the maximum inundated area was estimated using the different algorithms. Subsequently, the per-pixel occurrence probability \( p_{i,j} \) was estimated in the Google Earth Engine platform by Eq. 3.6:

\[
p_{i,j} = \frac{m_{ij}}{(n+1)}
\]

where \( n \) is the number of years on record (18, from 2000 until 2018), and \( m \) is the number of recorded occurrences of inundation during the period \( n \), being \( i \) and \( j \) the row and column index of each pixel in the image, respectively.

Return periods were calculated as the inverse of the occurrence probability to create inundated area frequency curves at the catchment scale. The same analysis was undertaken for streamflow data at the gauging stations located at Walgett and Bugilbone along the Namoi River, limiting the period of analysis to the number of years in which MODIS satellites have been in operation.

### 3.3. Results and discussion

#### 3.3.1. General comparison and performance

As an initial comparison, Figure 3.3 presents a log-scale boxplot of the inundated area distributions, calculated using the different algorithms.
The ISL algorithm estimated the highest maximum inundated area and an intermediate interquartile range (IQR) (37 km²), while BRAKC estimated the smallest mean inundated extent and the lowest IQR (12 km²). Given that a smoothing window was applied in the BRAKC algorithm (Nigro et al., 2014), the peak and variability of flood extents tended to decrease, which led to a narrower detection range of inundated areas. The OWL algorithm predicted the highest mean inundated areas and the highest IQR (56 km²), yet with the highest value lower than the ISL maximum. Given that the IQR is a resistant measure of data variability, and considering that flood events are mostly found above the third quartile in Figure 3.3, it can be inferred that the OWL algorithm could miss some flooded areas during large flood events, compared with the ISL algorithm. Whether this relates to over- or underestimation will be addressed later in this study by comparing the flood extents with “ground truth” inundations. The ensemble of the algorithms, on the right of the plot, resembles NDIS, yet with a lower IQR (15 km² compared with 38 km²), thereby suggesting that it predicts a smaller range of inundated extents. This behavior is a result of the ensemble technique, in which flooded pixels need to be classified as inundated in at least three of the five algorithms used, which leads to intermediate inundated extents. Table 3.3 presents the spatial performance of the different inundation algorithms using daily MODIS imagery for dates at which both satellites provided data (Landsat and MODIS Terra).

### Table 3.3. Evaluation of accuracy between Landsat water detection and different MODIS-based water detection algorithms.

<table>
<thead>
<tr>
<th>Date</th>
<th>Approach</th>
<th>Accuracy</th>
<th>Kappa</th>
<th>Commission*</th>
<th>Omission</th>
<th>Landsat flooded area (m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-12-12</td>
<td>NDIS</td>
<td>0.97</td>
<td>0.58</td>
<td>0.20</td>
<td>0.21</td>
<td>760,607,648</td>
</tr>
<tr>
<td></td>
<td>OWL</td>
<td>0.97</td>
<td>0.57</td>
<td>0.24</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ISL</td>
<td>0.97</td>
<td>0.60</td>
<td>0.20</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRAK</td>
<td>0.97</td>
<td>0.56</td>
<td>0.27</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRAKC</td>
<td>0.97</td>
<td>0.27</td>
<td>0.42</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENSEMBLE</td>
<td>0.97</td>
<td>0.59</td>
<td>0.23</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>2008-11-30</td>
<td>NDIS</td>
<td>0.99</td>
<td>0.50</td>
<td>0.18</td>
<td>0.30</td>
<td>114,383,101</td>
</tr>
<tr>
<td></td>
<td>OWL</td>
<td>0.97</td>
<td>0.57</td>
<td>0.24</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ISL</td>
<td>0.99</td>
<td>0.45</td>
<td>0.25</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRAK</td>
<td>0.99</td>
<td>0.36</td>
<td>0.35</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRAKC</td>
<td>0.99</td>
<td>0.01</td>
<td>0.50</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENSEMBLE</td>
<td>0.99</td>
<td>0.56</td>
<td>0.23</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>2008-12-16</td>
<td>NDIS</td>
<td>0.99</td>
<td>0.47</td>
<td>0.29</td>
<td>0.22</td>
<td>91,506,950</td>
</tr>
<tr>
<td></td>
<td>OWL</td>
<td>0.99</td>
<td>0.47</td>
<td>0.31</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ISL</td>
<td>0.99</td>
<td>0.54</td>
<td>0.24</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRAK</td>
<td>0.99</td>
<td>0.56</td>
<td>0.25</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRAKC</td>
<td>0.99</td>
<td>0.12</td>
<td>0.47</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENSEMBLE</td>
<td>0.99</td>
<td>0.49</td>
<td>0.30</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>2010-01-11</td>
<td>NDIS</td>
<td>0.99</td>
<td>0.63</td>
<td>0.19</td>
<td>0.18</td>
<td>246,838,512</td>
</tr>
<tr>
<td></td>
<td>OWL</td>
<td>0.99</td>
<td>0.62</td>
<td>0.20</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ISL</td>
<td>0.99</td>
<td>0.63</td>
<td>0.19</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRAK</td>
<td>0.99</td>
<td>0.61</td>
<td>0.23</td>
<td>0.15</td>
<td></td>
</tr>
</tbody>
</table>
The performance can be described as having poor, moderate, and substantial agreement for the different algorithms. The ISL algorithm performed best, presenting the highest kappa coefficients and lowest commission errors. This was followed by the ensemble of the algorithms. Comparing the rest of the algorithms, the performance was quite similar, except for the composite technique applied to the BRAK algorithm, which performed poorly. This may be related to the dynamic nature of flooding, where the inundation extent can change rapidly (within hours). Therefore, this is not captured in the three-day smoothing window of the BRAKC algorithm. Thus, while it has been shown that the composite technique (BRAKC) can generally assess the temporal dynamics of floods (Nigro et al., 2014), a spatial comparison with a snapshot of the inundation extents indicated poor performance. In the case of the ensemble, it performed slightly worse than ISL for single events. It has been recognized that the simplest multi-model average can outperform single models when modeling time series data by reducing the spread of the residuals (Ajami et al., 2006). However, because single events are evaluated in Table 3.3, the benefits of multi-model averaging may not arise.

Commission errors were generally higher than omission errors. The resolution difference between Landsat and MODIS (30 m and 500 m, respectively) introduced inaccuracies in the classification, which can affect both omission and commission errors, especially when mixed coverage classes are present, such as in this study (Boschetti et al., 2004). Further, the misclassification observed in the commission errors was caused by the difficulty in detecting flooded areas using the wavelength range of the bands contained in the surface reflectance product. Similarly, the omission errors were, on average, greater than 0.15, which was again a consequence of the different resolutions. For example, small flooded areas cannot
be captured at the MODIS resolution. Moreover, increased zenith angles at the moment of image capture may lead to failure to detect water surfaces (Ticehurst et al., 2014).

In general, there was a positive correlation between the kappa coefficients and the flood extent (correlation coefficients ranging from 0.42 to 0.72), while the commission and omission errors were logically inversely correlated with flood extent, with correlation coefficients averaging 0.52 and 0.64, respectively (not shown), meaning that the greater the inundated area in the catchment, the better the performance of the water detection algorithms. Thus, an increase of inundated extent improved the spatial performance because bigger inundated areas are more likely to be detected at the coarse resolution of MODIS, which leads to smaller omission errors. The same behavior was observed for the commission errors.

Water detection algorithms based on MODIS take advantage of the low reflectance of water, which also occurs with dark objects. When developing an algorithm for water detection based on MODIS imagery, Guerschman et al. (2011) observed that black clayey soils tended to be misclassified as water. The same clays occur across the western floodplain of the Namoi catchment (Green et al., 2011), and are periodically inundated during flood events. Black clayey soils in the catchment cover a large extent. For example, just the central black earth floodplains, which contain mainly black clayey soils, cover an area of over 3,478 km² (~ 8% of the catchment surface), and include only a fraction of the black clayey soils of the catchment (Welsh et al., 2014). Thus, during major flood events, misclassifications are reduced because these soils tend to be flooded, thereby contributing to reduced commission errors and improving the spatial performance. Some examples of the performance of the inundation detection are presented in Figure 3.A1 of the Supplementary Data.

3.3.2. Flood dynamics and relation with covariates

To evaluate the temporal prediction of the flood areas using the different approaches, the daily reflectance images were filtered to those with less than 25% of the area masked by clouds, shadows, or smoke from fires. The resulting flooded extents were plotted by selecting two years that presented large flood events, including daily discharges in the Namoi River at Walgett (Figure 3.4).
In general, the different approaches matched the discharge peaks recorded at the gauge station and were related to several of the major rainfall events. However, some discharge peaks also resulted in no peaks in the inundated areas (result not shown) and some flood detection occurred with no associated flood peak. Even though the MODIS imagery has a daily resolution, rain events and consequently discharge peaks are sometimes followed by several cloudy days, which precludes flood detection in some cases (Schumann and Moller, 2015), leading to some missed floods related to discharge peaks or heavy rain events.

Comparing the different approaches, the method with the highest noise was the BRAK algorithm. As a result, the ISL, OWL, NDIS, BRAKC, and ensemble technique detected the flood events more correctly in a time series and presented clear recession behavior. Figure 3.5 presents the relationship between the inundated areas using the different algorithms and the daily discharge at gauging stations at Walgett and Bugilbone.
Figure 3.5. Scatterplots between daily discharge and inundated areas obtained from the different algorithms at two gauge stations located in the Namoi River, at Walgett (a), and at Bugilbone (b), and their correlation coefficients. The dashed line is the linear fit for each algorithm.

The ensemble of the algorithms had a higher correlation with the river daily discharges, followed closely by the NDIS and ISL algorithms. In general, the correlation coefficients ($r$) in Figure 3.5, based on linear associations between the discharge and flood extents at both gauge stations, suggested reasonably strong correlations. In this case, the relationship between the discharge and inundated extent was approximated as linear, even though this is probably an oversimplification of the data, given the scatter of the points. In contrast, Ticehurst et al. (2014), on an area of 178 km$^2$ including the Victoria Lake in Australia, used second-order polynomial relationships. In the current study, the gauge stations and inundation occurred on flat areas of the catchment, where infiltration rates were low because of clay-dominated soils and open shrubland land cover, which may explain the different relationship.

The cross-correlation between daily discharge at different locations of the catchment and the daily inundated areas allowed identification of maximum cross-correlation coefficients and associated time lags. Figure 3.6 displays this relationship for the ISL algorithm, noting that the main flow direction in the catchment is westward.
Figure 3. 6. Maximum cross-correlation coefficients between daily discharges and daily inundation areas obtained from the ISL algorithm for different gauge stations (above) and their corresponding time lags (below).

Although the maximum cross-correlation values can be unrelated to the runoff-flood response, especially at high time lags, the moderate and strong associations between the variables, the relationship with the time lags, and the spatial relationship with other gauges (a similar general trend) and within the catchment suggest that these reflect the catchment runoff-flood behavior. In contrast, gauges with weak correlations between flood extent and discharge, and whose time lags differ from the surrounding gauges, must be interpreted as not representative of the overall catchment behavior. This can be observed in the gauges located just below the Chaffey Dam, which regulates downstream water flow, thereby affecting the natural streamflow pattern (Green et al., 2011).

In general, cross-correlation coefficients depend on the location of the stations in the catchment, with values ranging from 0 to 0.87. In general, downstream gauges, which receive greater amounts of the accumulated flow upstream in the catchment
and for longer periods, have higher cross-correlations. However, this general trend, which considers the flow path length, is also affected by the reach of the river. For instance, higher correlations occur in the Namoi River than in the Pian Creek.

The temporal response of floods, which can be interpreted as the time taken for flow waves to propagate downstream, shows a delay relative to the river discharge in the higher elevation (eastern) sectors of the catchment, with positive time lags of up to 15 days. This delay is reduced downstream, and, in the lower areas of the catchment, discharge peaks recorded at the stations occur after the maximum flood signal from the satellite (negative time lags). However, some areas have propagation times that differ from those observed immediately downstream, which are likely to be associated with local conditions affecting the flow (Hirpa et al., 2013). This time-lag difference between signals can help determine the main areas where floods would occur.

Although the downstream stations tended to be more representative of the general hydrological behavior of the catchment—which led to higher cross-correlation coefficients between daily discharges and daily flood extents—most had a delay in the flood–discharge response, which excluded any predictive capacity. Flood forecasting by using gauge stations requires selection of gauge stations within a catchment that have a close discharge–flood relationship (Turner, 2012) and where the discharge peak occurs some days before the flood response occurs. This creates an opportunity to investigate whether the inundation images may be able to provide information on downstream flooding conditions some days in advance.

The analysis undertaken enables the selection of stations that could be useful for flood forecasts. Stations with good cross-correlation values (light green dots in the cross-correlation values of Figure 3.6) and long time lags (blue dots in the time-lag values of Figure 3.6) are suitable to be included in an early warning flood forecasting system.

Given that rainfall is the main driver of floods, the rainfall–flood relationship was also assessed. Figure 3.7 displays a cross-correlation between daily lumped rainfall over the catchment and inundated extents for the different algorithms.
Chapter 3. Spatiotemporal evaluation of inundated areas

Figure 3.7. Cross-correlation between daily lumped rainfall over the catchment and flooded areas for the different algorithms.

According to Figure 3.7, the strongest correlation occurred consistently in all algorithms between 10 and 15 days. Other peaks in the correlation graph (Figure 3.7) could be caused by different types of runoff responses (such as infiltration excess, leading to correlation at lower time lags in the 40,000 km² catchment, or saturation excess, leading to correlations at longer time lags). However, they are more likely an artefact of the analysis based on catchment average rainfall. This removes the spatial variability of the rainfall, and shorter time-lag peaks could be related to locally concentrated heavy rainfall, while the longer time-lag peaks could relate to widespread rainfall and catchment flooding. We did not investigate this routing problem further here. Another way to evaluate this relationship is to examine the relationship between inundation extent and accumulated rainfall. This should detect temporal relationships between inundation areas from the algorithms for both the daily and eight-day composite images of the surface reflectance MODIS products (Figure 3.8).
Chapter 3. Spatiotemporal evaluation of inundated areas

Figure 3.8. Determination coefficients ($r^2$) from second-degree polynomial regressions for inundated area extents at increasing accumulated rainfall lags for daily MODIS images (a), and eight-day composite images (b).

Clearly, not all variation in the flood extents was explained by rainfall using a simple second-degree polynomial to describe the non-linear relationship between rainfall and flood extent. Other factors—such as soil properties, land cover, irrigation schedules, and the opportunity for water storage—may affect the surface water extent in the catchment (Pekel et al., 2016). In addition, antecedent soil moisture conditions also affect the catchment response, which indirectly highlights the relationship with the accumulated rainfall at different time lags.

In the case of the daily reflectance imagery, the highest $r^2$ related to NDIS and 23 days of accumulated rainfall, followed closely by the ensemble technique. Using the composite reflectance product, the $r^2$ reached a value greater than 0.6 with 13 days of accumulated rainfall. As was the case for the spatial performance and the assessment against daily discharges, the smoothing in the BRAKC algorithm led to lower correlations for daily rainfalls because it could miss the maximum flood extent.

The overall accumulated rainfall accounted for approximately 60% of the variation of the inundated extent. Using the composite reflectance imagery, the assessment of the flood dynamics improved, most likely because the output of the algorithms using the eight-day imagery was less noisy.

3.3.3. Frequency analysis

Comparing the different occurrence probability maps (Figure 3.9) indicated that the inundated probabilities in the GSW Landsat product were significantly smaller than those obtained from the MODIS imagery. This suggests that the different flood
extents were not just caused by the difference in the spatial resolution between the
satellites. More specifically, despite the longer period of operation, the lower
temporal resolution of the Landsat data compared with MODIS could lead to missing
flood events and under-prediction of flood extents (Mueller et al., 2016; Ticehurst
et al., 2014).

![Occurrence probabilities obtained from the recurrence layer of the GSW
product (above), the ISL algorithm (middle), and the OWL algorithm (below).]

**Figure 3.9.** Occurrence probabilities obtained from the recurrence layer of the GSW
product (above), the ISL algorithm (middle), and the OWL algorithm (below).

The flood frequency curves calculated from the probability maps (Figure 3.10)
differed from the standard flood frequency curves because the current study plotted
the flooded area, instead of discharge, on the Y-axis. The GSW recurrence layer
(derived from Landsat imagery) (Pekel et al., 2016) was used as a reference, noting
that the flooded areas were smaller than those based on MODIS imagery. Despite
the difference in behavior between the GSW product and MODIS imagery, the
general behavior of the GSW curve was similar to the curves using the ISL and OWL
algorithm below a return period of approximately two years. After this threshold, the
MODIS images had an inflection in the curves, leading to higher inundation areas. This inflection was probably caused by the topography of the catchment, which could also be observed in the river flood frequency data. In general, flood areas were slightly greater for all algorithms using daily reflectance imagery, compared with the eight-day composite images (data not shown), which confirms that larger temporal spacing can result in a loss of flood extent information.

Comparing the curves from the different algorithms indicated that, below return periods of 10 years, the inundation areas using the NDIS were considerably smaller than the others, yet all the MODIS flood detection approaches from the same reflectance product tended to have the same large return period threshold. Both the inundated area frequency curves derived from MODIS and the flood frequency curves derived from the gauging stations followed similar trends, as neither of them flattened out, even at large return periods of 20 years. This finding contrasts with the GSW product, which started to flatten out at approximately three years, thereby reducing its applicability in hydrological studies. Thus, an event with a return period of 10 years represents an inundated area of over 99.9% of the maximum inundated area of the GSW-derived frequency curve, while, with the MODIS-based algorithms, the same event represented, on average, 62% of the maximum inundation extent from the frequency curves. In the case of the river flood frequency curves, the same event for both stations accounted for an average of 65% of the maximum discharge obtained in the period studied, which was similar to the percentage estimated using MODIS imagery. This confirms that MODIS, despite its lower spatial resolution, is more useful than the Landsat imagery when analyzing processes that change rapidly, such as floods.
3.3.4. Final ranking of the algorithms

A qualitative ranking of the different algorithms (Table 3.4) summarizes the results of this study. The ranking is based on how well the algorithms performed against the benchmark Landsat image-derived flood areas, and how well they captured the flood dynamics relative to the gauge discharges. The algorithms that performed the best are ranked first, and the algorithms that performed the worst are ranked third.

Based on the performance against the flood occurrence from the Landsat images, and considering the mean kappa values (Table 3.3), algorithms ranked first had mean kappa coefficients greater than or equal to 0.55, algorithms ranked second had mean kappa coefficients between 0.5 and 0.55, and algorithms with mean kappa coefficients lower than 0.5 were ranked third. Then, based on the flood dynamics, and considering the lower limit of a strong association between variables (correlation coefficients > 0.7) and the mean correlation coefficients between inundated areas and discharges at the Walgett and Bugilbone stations (Figure 3.5), algorithms that had mean correlation coefficients greater than 0.7 were ranked first, mean correlation coefficients between 0.65 and 0.7 were ranked second, and mean correlation coefficients smaller than or equal to 0.65 were ranked third. The final rank was based on selecting the worst of the ranks previously mentioned.

Table 3.4. Qualitative ranking of MODIS-based water detection algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Spatial performance*</th>
<th>Flood dynamics</th>
<th>General</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISL</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ENSEMBLE</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NDIS</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>OWL</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>BRAK</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>BRAKC</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

*Ranking from 1 to 3, with 1 the best-ranked and 3 the worst-ranked algorithms.

Based on Table 3.4, the best methods, in both spatial performance and capturing the flood dynamics, were the ISL algorithm and ensemble technique. It is important to point out that, when evaluating the spatial performance, images with higher inundated areas should be considered, especially if these present low sensor zenith angles, because they represent more of the inundated areas, which might lead to lower uncertainty in flood modeling.
3.3.5. General discussion

A few previous studies compared mapping of water occurrence among different satellites (Ticehurst et al., 2015b). In general, two products were developed from Landsat imagery that calculate the inundation probability in time. One of these was used as reference in this study (Pekel et al., 2016). The other was developed for Australia and is available from the Geoscience Australia webpage (Mueller et al., 2016). Inundation probability maps from MODIS imagery have been generated at a smaller scale (Huang et al., 2014), but without the use of a Landsat reference and without development of frequency curves to assess the flood extent over time. To our knowledge, comparison of inundation probability frequencies between satellites has not previously been undertaken, and such a comparison increases understandings of the reliability of the different methods for flood studies. This study compared flood extents derived from MODIS and Landsat (GSW) at different spatial and temporal resolution, and highlighted the importance of high temporal resolution when studying dynamic processes, such as floods. However, this subject requires further analysis, especially for validating the inundation frequencies with ground measurements.

A comparison of the accuracy of different flood detection techniques using Landsat imageries was recently published (Munasinghe et al., 2018). However, no studies comparing MODIS-based algorithms for surface water detection have been performed. Moreover, most studies evaluating these algorithms using MODIS imagery have focused on assessing their spatial performance by using high spatial resolution images as a reference. The present study, which compared different MODIS-based algorithms, also considered flood dynamics by using daily discharges from gauge stations and total daily rainfall as references. In this case, the daily temporal resolution of MODIS imagery is able to capture some of the rapid changes occurring during flood events, which are frequently missed when using low temporal resolution datasets.

An advantage of capturing the dynamic nature of floods at a catchment scale is that it may help establish early warning flood forecasting systems by detecting gauging stations whose discharges present a strong correlation with flood extents and where discharge peaks are recorded some days before the floods occur. Similarly, by understanding how floods are spatially located and their recurrence, flood mitigation strategies can be implemented. Thus, the information derived from the present study can be used to direct flood prevention investment and to select design thresholds to prevent damage to public and private infrastructure. It can also support water resources planning by improving the quantification of changes in water volumes.
Future work may involve downscaling MODIS images to the spatial resolution of Landsat; however, an initial priority is reducing the computation requirements of the alternatives to achieve this. Further, synthetic aperture radar sources, which can be obtained from the Copernicus satellites, might allow greater precision in water detection and be used to improve the accuracy of the MODIS-derived flood extents via both validation and downscaling. However, because these data are only just coming available online, this will be a matter for future research.

### 3.4. Conclusions

Surface water can be detected using different MODIS-based algorithms. Based on the approaches used in this research, the ISL algorithm provided the highest spatial performance compared with the "ground truth" derived from Landsat images, and had the greatest inundation extent. When analyzing the temporal dynamic of floods, the best performance was obtained using the NDIS, the ISL, and the ensemble of algorithms. The BRAK algorithm created noisy estimates, with days with large inundation areas, but without matching discharges or rainfall. In contrast, the other algorithms captured the discharge peaks reasonably accurately, and even indicated distinct recession behavior.

The relationship between rainfall and flood extents based on composite images explained around 60% of the flood extent variation at a rainfall lag of 13 days. The relationship between stream discharge and the flooded area was spatially variable in the catchment, with generally higher correlations at gauges located downstream in the catchment. Cross-correlation coefficients reached up to 0.87, with a delay in recording the discharge peaks at the lowest stations of the catchment.

The GSW recurrence product captured significantly smaller inundated extents than did the MODIS-derived inundation, which can be explained by the lower temporal resolution of Landsat images. The GSW flood frequency curve was similar to the MODIS products for short return periods; however, at longer return periods, the inundated area derived from the MODIS imagery was considerably greater than the GSW estimates. In general, the daily imagery identified greater inundated areas than did the composite images because of the higher temporal resolution, which improved the estimates of the flood dynamics. The results of this study can reduce the uncertainty in flood volume estimates for real-time water resource planning and flood mitigation.
3.5. Appendix

Figure 3.A.1. Maps of water detection for three different dates with the results of Landsat 5 and the different MODIS-based algorithms, for a section in the downstream part of the catchment. Note that the Landsat image is missing some of the areas, while some of the approaches use different time windows. As a result, there are slight differences in the total area displayed between methods.

3.6. References


Chapter 3. Spatiotemporal evaluation of inundated areas


Chapter 3. Spatiotemporal evaluation of inundated areas


Geoscience Australia, 2015. Digital Elevation Model (DEM) of Australia derived from LiDAR 5 Meter Grid. Geoscience Australia, Canberra.


Hirabayashi, Yukiko; Mahendran, Roobavannan; Koirala, Sujan; Konoshima, Lisako; Yamazaki, Dai; Watanabe, Satoshi; Kim, Hyungjun; Kanae, Shinjiro, 2013. Global flood risk under climate change. Nature Climate Change 3(9), 816-821.


Chapter 3. Spatiotemporal evaluation of inundated areas


Scanlon, Bridget R; Reedy, Robert C; Faunt, Claudia C; Pool, Donald; Uhlman, Kristine, 2016. Corrigendum: Enhancing drought resilience with conjunctive use and managed aquifer recharge in California and Arizona. Environmental Research Letters 11(4), 49501.


101


Chapter 4

Comparison of surface water volume estimation methodologies that couple surface reflectance data and digital terrain models

Abstract

Uncertainty about global change requires alternatives to quantify the availability of water resources and their dynamics. A methodology based on different satellite imagery and surface elevation models to estimate surface water volumes would be useful to monitor flood events and reservoir storages. In this study, reservoirs with associated digital terrain models (DTM) and continuously monitored volumes were selected. The inundated extent was based on a supervised classification using surface reflectance in Landsat 5 images. To estimate associated water volumes, the DTMs were sampled at the perimeter of inundated areas and an inverse distance weighting interpolation was used to populate the water elevation inside the flooded polygons. The developed methodology (IDW) was compared against different published methodologies to estimate water volumes from digital elevation models, which assume either a flat water surface using the maximum elevation of inundated areas (Max), and a flat water surface using the median elevation of the perimeter of inundated areas (Median), or a tilted surface, where water elevations are based on an iterative focal maximum statistic with increasing window sizes (FwDET), and finally a tilted water surface obtained by replacing the focal maximum statistic from the FwDET methodology with a focal mean statistic (FwDET_mean). Volume estimates depend strongly on both water detection and the terrain model. The Max and the FwDET methodologies are highly affected by the water detection step, and the FwDET_mean methodology leads to lower volume estimates due to the iterative smoothing of elevations, which also tends to be computationally expensive for big areas. The Median and IDW methodologies outperform the rest of the methods, and IDW can be used for both reservoir and flood volume monitoring. Different sources of error can be observed, being systematic errors associated with the DTM acquisition time and the reported volumes, which for example fail to consider dynamic sedimentation processes taking place in reservoirs. Resolution effects account for a fraction of errors, being mainly caused by terrain curvature.
Statement of Contribution of Co-Authors

This chapter has been written as a journal article. The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
5. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for this publication is:


<table>
<thead>
<tr>
<th>Contributors</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ignacio Fuentes</td>
<td>Original idea; Coding; Data analysis; Writing</td>
</tr>
<tr>
<td>Jose Padarian</td>
<td>Edition; Suggestions</td>
</tr>
<tr>
<td>Floris van Ogtrop</td>
<td>Edition; Suggestions</td>
</tr>
<tr>
<td>R. Willem Vervoort</td>
<td>Supervision; Edition; Suggestions</td>
</tr>
</tbody>
</table>
4.1. Introduction

Changes in global climate and population result in increased uncertainty in relation to production and resource exploitation (Alcamo et al., 2000; Gleeson et al., 2012; Schewe et al., 2014). This is particularly relevant for water resources, whose availability and projections have been recently disputed (Zhang et al., 2016a; Scanlon et al., 2018). Given this uncertainty, several alternatives to quantify the availability of water resources must be developed to define water management plans or risk assessments with higher accuracy (Liuzzo et al., 2015; Poff et al., 2016).

Remote sensing is an important tool for studying surface water (Palmer et al., 2015; Schumann and Moller, 2015; Mueller et al., 2016). It has the advantage that it can be applied in conjunction with other direct measurements, and provides not only snapshots of ongoing processes, but can also capture the temporal fluctuation and seasonality of surface water processes (Doble et al., 2014; Siev et al., 2016).

The advantages of remote sensing data compared to other hydrological data lie in the opportunity to account for the spatial variability of processes (Mueller et al., 2016; Zhang et al., 2016b). Thus, while gauge and meteorological stations provide data for a specific location, satellite imagery reflect a larger spatial area (Gupta, 2018). Satellites also regularly pass over the same location, which provides a time series of the images, catching the temporal variability of some processes (Islam et al., 2010; Poff et al., 2016).

Most examples of water detection from space focus on the spatial extent of surface water, but not necessarily on the quantity (Islam et al., 2010; Zhou et al., 2017). In the cases where quantity is studied, they are generally coupled with gauge station measurements, bathymetry or digital terrain models (DTM), to estimate water volumes (Doble et al., 2014; Siev et al., 2016; Cohen et al., 2017; Scorzini et al., 2018). Several methods have been developed to obtain the depth of surface water. However, the performance of these methods has been mainly assessed against field measurements through individual events, rather than against the temporal dynamics of surface hydrological processes (Cohen et al., 2017). One example where temporal changes in hydrological processes have been taken into account, only covered short time periods (Siev et al., 2016).

Flood studies using remote sensing have been limited to short periods or single events due to challenges in acquiring detailed remote sensing information, which requires high computational storage capacity and due to the efforts involved in pre-processing imagery (Ma et al., 2015). However, several alternatives have recently been developed to cope with these tasks. One of the most important has been the development of the Google Earth Engine platform, which has multi-petabyte
processed and regularly updated geospatial datasets as well as a wide range of algorithms that facilitate spatial analysis and remote sensing functionality (Johansen et al., 2015; Gorelick et al., 2017).

A second main difficulty is to verify predicted inundation volumes, and this is because of lacking a frame of reference (Oreskes et al., 1994). Generally, the solution is to use water levels from gauging stations and dams or the estimation of the components of the water budget, which usually provide a rough estimate of the overall water availability (Alsdorf et al., 2007), but do not necessarily provide space and time verification. In addition, the scarcity of bathymetric continental data means that estimating surface water volumes in permanent water bodies is difficult (Zhang et al., 2016b). Moreover, irregularly inundated areas, in which water was absent when the DTMs were derived, are not consistently surveyed.

While the most common methodology for estimating surface water volumes assumes a flat water surface (Sievie et al., 2016), this is rarely the case. In flood processes for instance, the flow will be influenced by the topography of the terrain over which the water is passing. Even in big reservoirs and lakes, water surfaces are not totally flat. This may be caused by seiches or drain exits (Farhadzadeh et al., 2017). Therefore, assuming flat water surfaces for water volume estimations may lead to significative errors in water availability. The main objective of this study was to assess an automated methodology to estimate surface water volumes for flood events and reservoirs, taking into account that inundated areas are not necessarily flat. We then compared this methodology with two previously developed alternatives, which use surface reflectance imagery and elevation models as inputs. The final aim is to improve the ability to calculate flood time series using remote sensing data.

### 4.2. Materials and Methods

#### 4.2.1. Study areas and data

The study was carried out at nine locations that have digital terrain models (DTM) and where water volumes are continuously monitored. The first case study was the Menindee Lakes (Cawndilla, Menindee, and Pamamaroo), which are located in NSW, Australia (Figure 4.1). These lakes were modified in 1968 in order to increase the storage capacity and control floods in the Murray Darling basin (Burrel et al., 2013). A hot and dry climate on the floodplain depressions of the Lower Darling River characterizes the region where these lakes are located (Nicol, 2004). These characteristics in combination with regular floods lead to water fluctuations in the lakes, suitable for this analysis. Some characteristics of the Menindee lakes are shown in Table 4.1.
Other surface water bodies assessed were the Atoka reservoir, and the Ellsworth, Fork, Ray Roberts, Hubbard Creek, Tawakoni, and Stanley Draper lakes, located in Oklahoma and Texas in the United States (Figure 4.2 and Figure 4.3). The Oklahoma reservoirs are surrounded by smooth topographies within the Mississippi river basin and characterized by a humid subtropical climate (Rubel and Kottek, 2010).
The Texas reservoirs were selected based on the available information provided by the Texas Water Development Board webpage (https://www.twdb.texas.gov/). These lakes are located in northeast Texas, and also have a subhumid tropical climate (Figure 4.3). The primary purpose of these reservoirs is for water supply and conservation, but flood control is also an important purpose, at least for the Hubbard Creek and Ray Roberts lakes.
In the case of the Menindee lakes, the reference elevation data corresponds to the available Light Detection And Ranging (LiDAR) DTM at a 5 m resolution (Geoscience Australia, 2015). The campaign that obtained the LiDAR elevations covering the Menindee lakes was carried out in 2009, during a period in which all three reservoirs were empty, and therefore the observed elevation corresponds to the elevation of the bottom of the lakes (Figure 4.4).
Figure 4.4. Time series of volume storages for the Menindee lakes and LiDAR elevations. One GL is equivalent to $10^6 \text{ m}^3$.

For the Oklahoma lakes, bathymetric maps were obtained from the Oklahoma Water Resources Board with a resolution of 1.5 m (https://www.owrb.ok.gov/), which were resampled at 3 m and subsequently superimposed on the USGS National Elevation Dataset, which presents a 1/3 arc-second resolution (U.S. Geological Survey, 2019; Figure 4.5). However, the bathymetric maps were referenced to the specific normal elevation of each reservoir, instead of being referenced to the water elevation at which they were surveyed.

Figure 4.5. Terrain elevation of Oklahoma lakes.
In the case of Texas reservoirs, the bathymetric maps were obtained from the Texas Water Development Board webpage. The bathymetric studies were carried out in different surveys between 2008 and 2018 using multi-frequency sub-bottom profiling depth sounders (Solis et al., 2010; Solis et al., 2012a; Solis et al., 2012b; Leber et al., 2018). From elevation contour lines, bathymetric images were obtained at a 3 m resolution and superimposed on the USGS National Elevation Dataset (Figure 4.6).

![Figure 4.6. Terrain elevation of Texas lakes.](image)

Landsat 5 surface reflectance imagery was used to detect the surface water. The images were masked to remove clouds and cloud shadows. Images were used for the entire period of the operation of the satellite, but a filter was applied such that masked images which contained less than 99% of the reservoir extent were removed.
The recurrence layer of the Global Surface Water (GSW) Mapping Layers, v.1.0 from the Joint Research Centre (Pekel et al., 2016) was used as an input for the water detection using Landsat, available as a dataset in the Google Earth Engine platform.

### Table 4.1. Design characteristics of the studied reservoirs (Solis et al., 2010; Solis et al., 2012a; Solis et al., 2012b; Leber et al., 2018; Shivers, 2016).

<table>
<thead>
<tr>
<th>Reservoir</th>
<th>Storage capacity (GL)</th>
<th>Area (m²)</th>
<th>Maximum depth (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cawndilla lake</td>
<td>631.05</td>
<td>94,851,864</td>
<td>8.7</td>
</tr>
<tr>
<td>Menindee lake</td>
<td>629.49</td>
<td>163,936,661</td>
<td>8.1</td>
</tr>
<tr>
<td>Pamamaroo lake</td>
<td>277.73</td>
<td>66,861,857</td>
<td>7.8</td>
</tr>
<tr>
<td>Ellsworth lake</td>
<td>100.60</td>
<td>20,691,580</td>
<td>16.5</td>
</tr>
<tr>
<td>Stanley Draper lake</td>
<td>183.00</td>
<td>12,000,000</td>
<td>30.0</td>
</tr>
<tr>
<td>Atoka dam</td>
<td>152.00</td>
<td>23,000,000</td>
<td>18.3</td>
</tr>
<tr>
<td>Fork lake</td>
<td>785.11</td>
<td>108,816,018</td>
<td>18.29</td>
</tr>
<tr>
<td>Ray Roberts lake</td>
<td>972.59</td>
<td>115,926,351</td>
<td>32.31</td>
</tr>
<tr>
<td>Hubbard Creek lake</td>
<td>392.46</td>
<td>63,483,092</td>
<td>18.29</td>
</tr>
<tr>
<td>Tawakoni lake</td>
<td>1,075.22</td>
<td>151,049,049</td>
<td>19.23</td>
</tr>
</tbody>
</table>

#### 4.2.2. Water detection

Munasinghe et al. (2018) argue that the best performance for water detection methods on Landsat imagery is obtained using supervised classification, rather than using normalized indices. Therefore, a classification and regression tree analysis (CART) was applied to the reflectance bands of the Landsat imagery to delineate inundated areas by selecting known surface water and dry land end-members. This was done by drawing polygons classified as water on several images, previously masked to remove clouds and cloud shadows. Other polygons were also delineated on dry areas with different land covers, and classified as dry polygons, which produced surface water classified images (Figure 4.7).
Figure 4. 7. Two examples of areas where polygons classified as water (blue polygons) and dry (yellow polygons) were drawn to feed the classification and regression tree analysis (CART) classifier with their corresponding classifications. The background images where the polygons were drawn correspond to color infrared Landsat images.

A total of 188 and 205 polygons in surface water and dry land areas were used to train the classifier, respectively, using 15 different Landsat 5 images covering Australia and the United States on different dates. Nevertheless, since the classifier is fed by pixels rather than polygons, the classification was carried out using 56,438 dry land and 133,194 surface water pixels.

Due to the difficulties caused by topography and dark lithologies, in which mountain shadows and dark lithologies tend to be classified as water due to the low reflectance, the recurrence layer of the GSW dataset (Pekel et al., 2016) was also used as an input for the classification. This was appended as a band to the surface reflectance bands of the Landsat images and passed to the CART classifier.

### 4.2.3. Water depths

Once the inundated areas in the images were delineated, different methods were used to estimate the water volumes. This involves using the DTM products, including
bathymetries, to obtain the depth of water at the surface, which was subsequently multiplied by the area of inundated pixels.

The first method (Max) assumes that the water surface during floods is flat, based on Siev et al. (2016) who studied the seasonal change in water volumes in a floodplain of almost 5000 km² in Cambodia. In it, several polygons with inundated areas are overlain by a DTM, and the maximum elevation of water in those polygons is assumed to be the elevation of the surface water. Subsequently, the DTM is subtracted from these elevations in each polygon to get the water depth.

Since the Max methodology may be strongly affected by errors in the surface water classification, it was complemented by another hydro-flattening methodology, subsequently referred to as “Median”. It consisted of a line vectorization of the perimeter of inundated areas, which was then buffered 2.5 m at each side. Subsequently, the DTM was clipped by the buffered layer extent and the median contour DTM elevation was estimated and extrapolated for each inundated area. Finally, the DTM was subtracted from the median elevation within the perimeter of inundated areas to get the water depths.

The third method (FwDET) was developed by Cohen et al. (2017) for flood analysis. It involves the conversion of inundated areas into polygons to obtain the elevations at the perimeter of polygons. Subsequently, it applies a focal statistic (focal maximum) in a series of iterations with increasing window sizes to populate the area inside the polygons with water elevations. The final stage involves subtracting the water elevations from the original DTM to get the water depths. Negative water depths are converted to 0, and a final low-pass filter with a kernel of 3 pixels is used to smooth any abrupt change in the water elevations. An important detail is that the number of iterations corresponds to the minimum number of iterations needed to completely populate all the inundation polygons. Additionally, a modification of the FwDET methodology was implemented (subsequently referred to as FwDET_mean), which replaced the focal maximum statistic of the original methodology with a focal mean statistic.

The last method, inverse distance weighting (IDW), also corresponds to a modification of the FwDET algorithm to improve volume estimates in water reservoirs, because the Cohen et al. (2017) study reports methodological errors in the estimation of reservoir volumes. The new methodology consists of delineating the perimeter of inundated areas and applies a random sampling of the perimeter elevations using a buffer of 2.5 m on each side of the perimeter contour, setting the number of sampling points to 5000. Then, an inverse distance weighting interpolation is applied to the sampled points to obtain the elevation of the water, which is subsequently subtracted from the DTM to obtain the water depth (Figure 4.8).
A filter was applied to all methods, such that polygons with less than six inundated pixels were removed from the Landsat images.

All preprocessing steps, the water detection, and the different methodologies were implemented in Google Earth Engine.

### 4.2.4. Covariates and performance of methodologies

In the case of Cawndilla lake, an additional 1 m resolution LiDAR DTM was also obtained to assess how the resolution of the DTM affects the volume estimations. Thus, for the Cawndilla and the Hubbard lakes (which already have 1 m DTMs), the original resolution was progressively reduced to 3, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 1000, 2000, and 3000 m. The analysis of the slope and its effect on the estimates compared the slope of both reservoirs and the performance of the volume estimates at the different resolutions.

The final volume estimates were compared against volumes reported by government organizations from gauging station data. In the case of the Menindee lakes, reported
daily volumes were obtained from the Water NSW webpage (https://realtimedata.waternsw.com.au/), whilst daily volumes from the Oklahoma and Texas reservoirs were obtained from the USGS platform (https://waterdata.usgs.gov/nwis). In the case of the USGS datasets, only volumes that were approved for publication were used. Reported volumes are obtained based on ratings tables generated from initial bathymetric surveys, which provide the relationship between storage levels and dam volumes.

As a performance evaluation the root mean square error and the bias of the relationship between observed and estimated volumes were estimated, in addition to the coefficient of determination and a p-value for the linear regression between both datasets.

4.3. Results

4.3.1. Menindee lakes

As the LiDAR DTM used in the volume estimates of the Menindee lakes was developed when no water was stored, quantifying water stored in the reservoirs does not require further processing of the terrain model. In addition, the DTM can be extrapolated to quantify water volumes associated with flood events in areas that are dry most of the time. A summary of the performance of the different methodologies for the Menindee lakes is presented in Table 4.2.
Table 4.2. Performance of the different methodologies used to estimate water volumes in the Menindee lakes.

<table>
<thead>
<tr>
<th>Reservoir</th>
<th>Method</th>
<th>$p$-value</th>
<th>$R^2$</th>
<th>RMSE (GL or m$^3$ x 10$^6$)</th>
<th>bias (GL or m$^3$ x 10$^6$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cawndilla</td>
<td>Max</td>
<td>0.0</td>
<td>0.96</td>
<td>88.25</td>
<td>40.65</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0</td>
<td>0.99</td>
<td>85.85</td>
<td>-53.91</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>0.0</td>
<td>0.98</td>
<td>56.27</td>
<td>21.75</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>0.0</td>
<td>0.98</td>
<td>174.69</td>
<td>-111.59</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>0.0</td>
<td>0.99</td>
<td>81.86</td>
<td>-49.32</td>
</tr>
<tr>
<td>Menindee</td>
<td>Max</td>
<td>0.0</td>
<td>0.95</td>
<td>254.59</td>
<td>143.30</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0</td>
<td>0.99</td>
<td>60.96</td>
<td>-36.92</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>0.0</td>
<td>0.97</td>
<td>250.22</td>
<td>129.17</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>0.0</td>
<td>0.98</td>
<td>164.03</td>
<td>-105.79</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>0.0</td>
<td>0.99</td>
<td>59.24</td>
<td>-36.74</td>
</tr>
<tr>
<td>Pamamaroo</td>
<td>Max</td>
<td>0.0</td>
<td>0.80</td>
<td>141.79</td>
<td>107.47</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0</td>
<td>0.97</td>
<td>48.62</td>
<td>-41.28</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>0.0</td>
<td>0.85</td>
<td>113.63</td>
<td>93.63</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>0.0</td>
<td>0.89</td>
<td>105.31</td>
<td>-94.98</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>0.0</td>
<td>0.96</td>
<td>48.72</td>
<td>-41.18</td>
</tr>
</tbody>
</table>

Except for the Cawndilla lake, the Max method tends to have higher noise compared to the other methods and tends to overestimate the volumes. This methodology appears more sensitive to the water detection technique. While other methodologies use the entire perimeter of the surface water, which offsets errors generated in the delineation of surface water, the maximum elevation methodology, by picking up only one value of elevation and extrapolating this to the entire water surface, is prone to errors at the water detection step.

The FwDET methodology leads to high $R^2$ coefficients, but also to high errors in the Menindee and Pamamaroo lakes. In this case, the bias is positive due to a propagation of the maximum elevations in the successive iterations of the focal statistic, and therefore propagates the errors caused in the water detection and by the mismatch between the resolutions of the DTM and the surface reflectance product. The FwDET_mean, the Median, and the inverse distance weighting interpolation estimates result in a negative bias, which was higher for the first method (Table 4.2). When using the FwDET_mean method, increasing the iterations, which increased the window size in the focal mean statistic, tends to smooth and diminish the volumes. Therefore, the minimum number of iterations was based on the number of iterations needed to fill the entire area inside the inundated polygons (Cohen et al., 2017). However, despite this, considerable negative bias is introduced. Another disadvantage of this methodology is that the focal statistic iteration with increasing window sizes is computationally expensive for big areas at high resolutions, which constrains its use.
Chapter 4. Comparison of surface water volume estimation methodologies

Using both the Median and the inverse distance weighting interpolation methodology decreases the errors and the bias, improving the volume estimates. By assessing the residuals between estimated and observed values using the IDW method it is clear that the negative bias increases at higher observed volumes (Figure 4.9b).

![Figure 4.9](image)

**Figure 4.9.** Water volume estimates (a), residuals (b) and area-volume relationship (c) obtained using the inverse distance weighting interpolation methodology on the Cawndilla, Menindee and Pamamaroo lakes. The cumulative frequency of slopes in the reservoirs is also presented (d).

From the analysis, a relationship between inundated areas and volumes can be derived (Figure 4.9c), which facilitates further analysis of lake volume by just fitting a regression curve to the data. The shape of the curves also identifies characteristics of the reservoirs. In this case, the Menindee lakes have an exponential increase in volumes with respect to the inundated area, which implies small slopes inside the lakes, but a sharp slope at the perimeter of the lakes, which can also be inferred from the circular shape of the lakes in Figure 4.4.

The area/volume relationship indicates a steeper slope for greater inundated areas. The slope is lower at the Cawndilla lake and is confirmed by the cumulative frequency of slopes from the lakes (Figure 4.9d).
4.3.2. Oklahoma reservoirs

The Oklahoma lakes have significantly smaller storage volumes, and as a result have generally smaller root mean square errors (RMSEs) than the Menindee lakes, despite their lower determination coefficients (Table 4.3). However, all reservoirs have a systematic bias, which was always negative for the FwDET_mean and positive for the Max and FwDET methodologies over the entire range of observations. In the case of the Median and IDW methods, the bias is both negative and positive, and as in the Menindee lakes, they have the lower RMSE values.

Table 4.3. Performance of the different methodologies used to estimate water volumes in Oklahoma reservoirs.

<table>
<thead>
<tr>
<th>Reservoir</th>
<th>Method</th>
<th>p-value</th>
<th>$R^2$</th>
<th>RMSE (GL or m$^3 \times 10^6$)</th>
<th>bias (GL or m$^3 \times 10^6$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atoka dam</td>
<td>Max</td>
<td>0.0</td>
<td>0.25</td>
<td>150.01</td>
<td>133.05</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0</td>
<td>0.93</td>
<td>13.83</td>
<td>-10.99</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>0.0</td>
<td>0.28</td>
<td>98.03</td>
<td>68.03</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>0.0</td>
<td>0.77</td>
<td>68.61</td>
<td>-67.12</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>0.0</td>
<td>0.77</td>
<td>24.79</td>
<td>-21.81</td>
</tr>
<tr>
<td>Ellsworth lake</td>
<td>Max</td>
<td>0.0</td>
<td>0.79</td>
<td>68.85</td>
<td>67.63</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0</td>
<td>0.95</td>
<td>11.07</td>
<td>9.99</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>0.0</td>
<td>0.81</td>
<td>40.77</td>
<td>39.12</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>0.0</td>
<td>0.95</td>
<td>14.46</td>
<td>-12.13</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>0.0</td>
<td>0.94</td>
<td>13.30</td>
<td>12.33</td>
</tr>
<tr>
<td>Stanley Draper</td>
<td>Max</td>
<td>0.0</td>
<td>0.84</td>
<td>28.11</td>
<td>26.69</td>
</tr>
<tr>
<td>lake</td>
<td>Median</td>
<td>0.0</td>
<td>0.96</td>
<td>5.24</td>
<td>-4.36</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>0.0</td>
<td>0.79</td>
<td>20.46</td>
<td>17.85</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>0.0</td>
<td>0.97</td>
<td>24.16</td>
<td>-24.04</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>0.0</td>
<td>0.96</td>
<td>13.17</td>
<td>-12.87</td>
</tr>
</tbody>
</table>

The bias can be explained by the fact that the different bathymetric maps were referenced to the mean water level elevation of the reservoir. This choice is generally made because the surveys usually take several days to complete, and can be finished even in different seasons. However, as the reference elevation does not correspond to the reservoir level at the moment of the survey, it introduces consistent errors throughout the entire range of observations. One solution to this problem might be to use the mean elevation of the surveyed days rather than the mean elevation of the reservoir for the generation of the bathymetry maps, especially if the variation of water depths during the survey is lower than the variation in the entire reservoir monitoring period.
Additionally, this bias might be simply removed from the estimates assuming that it is caused by using the mean level elevation of the reservoir instead of the water level at the moment of the survey. The residuals in the Atoka reservoir showed an abrupt drop at the higher end of the volume observations (Figure 4.10b). This change is related to the change of DTM from the 3 m resolution bathymetry map to the 1/3 arc-second resolution USGS National Elevation Dataset.

![Figure 4.10](image)

**Figure 4.10.** Water volume estimations (a), their residuals (b), and the area-volume relationship (c) obtained using the inverse weighted distance interpolation methodology on the Oklahoma reservoirs. The cumulative frequency of slopes in the reservoirs is also presented (d).

Analyzing the area–volume relationship (Figure 4.10c) suggests quite a different behavior compared to the Menindee lakes, as the relationship is almost linear. In this case, the shape of the reservoirs also differs from the circular shape observed in the Menindee lakes, and has less steep slopes at the edges. The steepest slopes can be observed in the Stanley Draper reservoir and the smoothest bathymetry in the Ellsworth lake (Figure 4.10d).
Chapter 4. Comparison of surface water volume estimation methodologies

4.3.3. Texas reservoirs

The performance of the different methods for the Texas reservoirs is in Table 4.4. All relationships are highly correlated with the observed data. In general, both the Max and the FwDET methodologies had a positive bias compared to the others.

Table 4.4. Performance of the different methodologies used to estimate water volumes in Texas reservoirs.

<table>
<thead>
<tr>
<th>Reservoir</th>
<th>Method</th>
<th>$p$-value</th>
<th>$R^2$</th>
<th>RMSE (GL or m$^3$ x 10$^6$)</th>
<th>bias (GL or m$^3$ x 10$^6$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hubbard Creek lake</td>
<td>Max</td>
<td>$8.10 \times 10^{-15}$</td>
<td>0.90</td>
<td>86.71</td>
<td>74.68</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>$1.90 \times 10^{-23}$</td>
<td>0.98</td>
<td>22.22</td>
<td>-12.12</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>$7.80 \times 10^{-20}$</td>
<td>0.96</td>
<td>77.15</td>
<td>68.78</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>$6.14 \times 10^{-11}$</td>
<td>0.96</td>
<td>38.89</td>
<td>-33.56</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>$5.24 \times 10^{-22}$</td>
<td>0.97</td>
<td>27.14</td>
<td>-19.02</td>
</tr>
<tr>
<td>Tawakoni lake</td>
<td>Max</td>
<td>$4.19 \times 10^{-15}$</td>
<td>0.63</td>
<td>288.35</td>
<td>262.16</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>$8.34 \times 10^{-25}$</td>
<td>0.82</td>
<td>85.49</td>
<td>-63.75</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>$1.47 \times 10^{-23}$</td>
<td>0.80</td>
<td>263.63</td>
<td>219.52</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>$2.43 \times 10^{-31}$</td>
<td>0.89</td>
<td>175.90</td>
<td>-166.11</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>$4.21 \times 10^{-30}$</td>
<td>0.88</td>
<td>119.07</td>
<td>-108.63</td>
</tr>
<tr>
<td>Ray Roberts lake</td>
<td>Max</td>
<td>$2.51 \times 10^{-47}$</td>
<td>0.90</td>
<td>188.12</td>
<td>162.02</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>$4.26 \times 10^{-49}$</td>
<td>0.91</td>
<td>148.14</td>
<td>-90.13</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>$3.98 \times 10^{-47}$</td>
<td>0.90</td>
<td>172.35</td>
<td>145.66</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>$1.59 \times 10^{-42}$</td>
<td>0.87</td>
<td>211.32</td>
<td>-166.59</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>$1.15 \times 10^{-54}$</td>
<td>0.93</td>
<td>166.04</td>
<td>-121.30</td>
</tr>
<tr>
<td>Fork lake</td>
<td>Max</td>
<td>$1.79 \times 10^{-04}$</td>
<td>0.18</td>
<td>140.44</td>
<td>183.77</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>$1.39 \times 10^{-11}$</td>
<td>0.48</td>
<td>61.29</td>
<td>-50.81</td>
</tr>
<tr>
<td></td>
<td>FwDET</td>
<td>$1.03 \times 10^{-12}$</td>
<td>0.51</td>
<td>125.51</td>
<td>116.94</td>
</tr>
<tr>
<td></td>
<td>FwDET_mean</td>
<td>$3.37 \times 10^{-14}$</td>
<td>0.57</td>
<td>126.39</td>
<td>-122.41</td>
</tr>
<tr>
<td></td>
<td>IDW</td>
<td>$3.10 \times 10^{-11}$</td>
<td>0.47</td>
<td>81.78</td>
<td>-74.00</td>
</tr>
</tbody>
</table>

Again, the Median and IDW methodologies outperform the rest of the methodologies. Comparing the results with the other lakes, it can be observed that there are also increased negative errors for greater storage volumes (Figure 4.11b). A greater reservoir storage volume relates to a greater flooded area. Since the perimeter of reservoirs in surface reflectance images is often composed of mixed land surfaces (flooded and dry land), these areas are more susceptible to water detection classification errors. These errors and the mismatch between the resolution of DTMs and the Landsat images can cause errors in the elevations at the perimeter of the reservoirs used to fill the water elevations. This results in bigger errors of volume estimates associated with greater inundated areas.
Chapter 4. Comparison of surface water volume estimation methodologies

Figure 4.11. Water volume estimations (a), their residuals (b), and the area-volume relationship (c) obtained using the inverse distance weighting interpolation methodology on the Texas reservoirs. The cumulative frequency of slopes in the reservoirs are also presented (d).

The relationship between inundated areas and reservoir volumes is again fairly good. In this case the smoothest bathymetry is in the Tawakoni lake (Figure 4.11d) despite having the highest storage capacity.

A comparison of water depths for the Hubbard Creek lake is shown in Figure 4.12 for 15 February 1991. It can be observed that there is a clear difference between the Max and the FwDET methodologies and the rest. The Max and the FwDET lead to greater water depths, which may be the result of a combination of causes: mixed land covers within surface reflectance image pixels, a mismatch between resolutions of the DTM and the surface reflectance product, and errors in the water detection step. The best performers, the Median and IDW methodologies, give similar water depth maps.
4.3.4. Resolution and slope

The decrease in the resolution (greater pixel size) leads to a decrease in the RMSE at very small pixel sizes. However, this decrease is greatest at a greater pixel resolution in lake Cawndilla (around 300 m; Figure 4.13a) compared to the Hubbard reservoir (50 m; Figure 4.13b). After the initial RMSE decrease, a steady increase in the RMSE is observed in both reservoirs, with a smaller increase for Hubbard lake. Although the overall RMSE is greater at Cawndilla, if the errors are normalized to the initial RMSE, the prediction error at Cawndilla lake is around two times the initial error, while at Hubbard Creek lake errors are about four times greater than the RMSE at the lowest pixel size. However, determination coefficients decrease faster in the Cawndilla lake, thus showing a faster deterioration of the precision of the volume prediction with pixel size.
Chapter 4. Comparison of surface water volume estimation methodologies

Figure 4.13. Effect of pixel size on the volume estimation errors for the Cawndilla (a) and the Hubbard (b) reservoirs. The slope differences between reservoirs are also presented (c).

The cumulative slope frequency distribution for the Cawndilla and Hubbard reservoirs is presented in Figure 4.13c. The slopes in the Cawndilla lake are much lower than the Hubbard reservoir, which has a much greater range of slopes. The design characteristics of both reservoirs indicate that the maximum depth of the Hubbard Creek reservoir is more than two times the depth of Cawndilla lake. This leads to the slightly different behaviors obtained when analyzing the resolution effect on the performance of volume estimations.

4.4. Discussion

Several alternatives have been developed to deal with problems in estimating surface water volumes. Sawunyama et al. (2006) found a power relationship between areas and volumes in small reservoirs within a catchment in South Africa, and Liebe et al. (2005) estimated volumes of small reservoirs in Ghana based on
their surface areas coupled with bathymetric data. While such estimates can be useful, extrapolation of the results to different reservoirs should be avoided because, as can be seen in this study, reservoirs have different area–volume relationships based on their particular bathymetries.

Regarding the pixel size effect on volume estimates, there have been several studies evaluating the horizontal resolution and its impact on hydrologic processes, with quite different results. Usery et al. (2004), studying topographic indices, such as the topographic wetness index (TWI), which depends on the slope, concluded that between 3 m and 30 m pixel sizes the results were unaffected, while a further reduction in resolution led to a degradation. In contrast, Sørensen and Seibert (2007) found a considerable degradation in indices, moving from a 5 m to a 10 m pixel size, but Cai and Wang (2006) did not report any worsening by diminishing the resolution from 30 m to 90 m. With respect to volumetric studies, Walczak et al. (2016) showed that a decrease from 1 m to 100 m resolution using LiDAR images led to approximately a 10% change in polder volumes. Additionally, the same study concluded that a 10% change in resolution led to little impact in the results. In the current study, the change of resolution in the flat areas of the terrain using the IDW methodology did not affect the volume estimates for a pixel size smaller than 300 m. There may even be a slight increase in performance as the resolution decreased from 1–3 m up to 300 m. Thus, despite the high DTM resolution, the accuracy of the methodology is constrained by the data used in the water detection, since there is a mismatch in the resolutions of the surface reflectance and the DTM data. For example, even if a 1 m resolution DTM with high vertical accuracy is used, the Landsat imagery used in the water detection constrains the water perimeter to a resolution of 30 m. This affects the volume estimates as elevations can suddenly change in an interval of 30 m, which is one of the reasons that might explain the increase in the performance when increasing the pixel size (Figure 4.13). However, this behavior may also be explained by the flat surrounding terrain, and it could be different for steeper topographies surrounding lakes and reservoirs. Additionally, the 30 m resolution used when delineating the perimeter elevations causes considerably higher methodological errors in water volume estimates. In particular, this occurs when using methods that take extreme values in the distribution of elevations at the reservoir perimeters, such as the Max and FwDET methodologies, which lead to significant positive biases.

One of the difficulties analyzing the effect of the topography and the resolution of images on the volume estimates is that these cannot be taken into account in isolation because the pixel size affects the elevation on a pixel basis. In order to understand how slope affects area (and volume) estimates a diagram is presented (Figure 4.14). Constant slopes (Figure 4.14a) will offset gains (blue areas) and losses (red areas) in inundation area and volume at any pixel scale, independently of the slope angle. However for curved slopes, changes in the slope over distance
(the second derivative of the elevation, or curvature; Figure 4.14b) cause biased errors in the volume estimates. In this last case, the increase in the pixel size and the higher curvature affect the errors in area and volume estimates by increasing the differences between the gains and losses in inundation area.

![Figure 4.14](image.png)

**Figure 4.14.** Area errors produced by pixel representation (dashed lines) of actual terrain elevation (continuous lines) for steady (a) and curved (b) slopes. Blue and red areas represent area gains and losses, respectively.

In this study, the curvature of the Cawndilla lake ranges between $-0.22$ and $0.23$ with a standard deviation of $0.007$, while the range and standard deviation of curvatures are significantly higher in the Hubbard Creek lake (from $-1.13$ to $1.21$ and $0.05$, respectively). This may explain the higher impact of increasing the pixel size on volume estimates at Hubbard Creek lake.

While the Median and the IDW methodologies are clearly superior, there is a clear and consistent negative bias in both. This systematic error may be a result of the digital terrain models, which were mostly acquired in the last 10 years, while the reference volumes (for example, calculated from original bathymetric surveys) date from 1987. In most cases, reservoir volumes reported are related to terrain conditions prior to the start of the operation of the reservoirs, a moment in which the terrain elevation was known. In lakes, these are generally linked to the date of the generation of a bathymetry, from which rating or elevation–area–volume curves are developed and usually associated with a gauge or limnimeter for continuous monitoring. However, sedimentation processes are continuously taking place at different rates, affecting the reservoir bathymetries, and this is generally not considered when monitoring the reservoir volumes (Palmieri et al., 2001), which implies an intrinsic error associated with the reference data. For example, the bathymetry used for the Hubbard Creek reservoir was carried out in 2018. It complemented a bathymetric survey carried out in 1997. The sedimentation processes between both surveys result in an estimated loss rate of storage capacity
of 0.68 GL y\(^{-1}\), which in 20 years accounts for a loss of 13.66 GL (Leber et al., 2018).
In the case of Fork lake, it has sedimentation rates between 1.46 and 2.32 GL y\(^{-1}\) (Solis et al., 2012), which results in 46.72 and 74.24 GL of lost capacity since the beginning of operations. However, these rates can hardly be applied continuously because they represent mean rates in the period of analysis, and sedimentation can abruptly change depending on rare climatic events.

In this study, since most of the bathymetries were surveyed in the last 10 years, a projection of the terrain to estimate past volumes must lead to lower estimates than the reported values (errors), due to lost reservoir storage capacity through sedimentation. However, the reported values based on limnimeters or gauges in the reservoirs are also over predicting the reservoir volumes because these are based on a projection of past conditions of the terrain. This partially explains the systematic negative errors obtained in the different reservoirs.

A way to deal with changing landscapes and bathymetries is to use terrain models obtained at higher frequency intervals. While such tasks may be unrealistic at a global scale, some alternatives that use remote sensing have been applied. For example, Zhang et al. (2016c), by using TanDEM-X imagery, were able to obtain bathymetric maps of several lakes in Brazil. Such an approach, if repeatedly carried out in time, can lead to a better knowledge of sedimentation rates and changes in landscape and water storage capacities.

Other sources of error, include the accuracy of the DTM sources. For instance, the LiDAR DTM used specifies a vertical accuracy of at least \(\pm 0.3\) m and a horizontal accuracy of at least 0.8 m (Geoscience Australia, 2015). Even though no accuracies are reported in the bathymetric studies used, the USGS (Wilson and Richards, 2006) reported vertical accuracies of 0.2, 0.28, and 0.46 m at the 95 percent confidence interval in the surveyed data, bathymetric model, and data contour map generated through echo sounder bathymetric studies, respectively. Therefore, even though the errors in high resolution DTM are low, these still may lead to large volume errors in big reservoirs.

Since terrain models and surface reflectance data at high spatial resolutions are progressively more available at global scales, these allow the monitoring of reservoirs and floods. While both the IDW and the Median methodologies had better performances than the rest, floods occurring in floodplain areas are usually taking place along river channels, which implies an elevation gradient. This precludes the use of techniques that assume a flat water surface, such as the Median methodology. Therefore, for such instances, the alternative IDW methodology might be a better choice for flood water volume estimates, which might be useful for water management plans and risk management studies. However, these estimates might have a stronger basis if the temporal resolution and the operation times of the
datasets allow for continuous and long-term monitoring, which might also allow analyzing the temporal dynamic of such events (Fuentes et al., 2019).

While other approaches to study surface water dynamics have been developed by coupling satellite imagery and altimetry data (Duan and Bastiaanssen, 2013; Busker et al., 2019), these methods only provide data on fluctuations in volumes and not total volumes, which constrains their applicability for water management planning.

### 4.5. Conclusions

The Max and the FwDET methodologies resulted in considerable errors and high bias in volume estimation. Both methods were strongly affected by a combination of issues, including: mixed land covers within surface reflectance image pixels, a mismatch between DTM and surface reflectance product resolutions, and errors in the water detection step. The FwDET_mean method had an intermediate performance. Both the Median and IDW methodologies outperformed the rest across the studied reservoirs. However, a negative bias was systematically observed in the estimates.

Pixel size and the curvature of the terrain were common factors introducing errors that affect area and volume estimations. Other sources of systematic errors are associated with the terrain models and the reported volumes stored in reservoirs. These last fail to consider the bathymetric changes occurring in reservoirs due to sedimentation processes, which can lead to an overestimation of water availability.

Even though some relationships between reservoir areas and volumes have been discussed in the literature, the extrapolation of such results to different scenarios must be avoided since area–volume relationships are specific to each reservoir and depend on their specific bathymetries.

Future research using higher resolution imagery or processed datasets for water detection, including synthetic aperture radar, could be used to improve the results. Moreover, the Surface Water Ocean Topography mission, projected for 2021, will allow a more thorough understanding of surface waters, but different alternatives will still be required to study the dynamics of hydrological processes and their recurrence since decades of available data will still be required to study long term hydrologic processes.
4.6. References


Geoscience Australia, 2015. Digital Elevation Model (DEM) of Australia Derived from LiDAR 5 Meter Grid; Geoscience Australia: Canberra, Australia.


Leber, N., Holmquist, H., Iqbal, K., Duty, J., 2018. Volumetric Survey of Hubbard Creek Reservoir; Texas Water Development Board: Austin, TX, USA.


Chapter 4. Comparison of surface water volume estimation methodologies


Shivers, M., 2016. Dam breach study and inundation mapping of eleven dams owned by Oklahoma City, Oklahoma. Master’s Thesis, Oklahoma State University, Stillwater, OK, USA.


Solis, R., Kemp, J., Connell, T., Weyant, H., McEwen, T., Brock, N., 2012b. Volumetric and Sedimentation Survey of Lake Fork Reservoir; Texas Water Development Board: Austin, TX, USA.


132
Chapter 4. Comparison of surface water volume estimation methodologies


Chapter 5

Comparing volume estimates from flood detection algorithms for spatiotemporal analysis of floods

Abstract

Episodic flood events have a significant impact on ecosystems and human settlements and can be important drivers for the water availability. This study maps the flood extension at a catchment scale, and determines the volumes associated with inundation events for selected return periods by coupling a water detection algorithm and three methods for water depth estimation. The study was carried out in the Namoi catchment of Australia by using the Google Earth Engine platform. The extension of inundated areas was obtained by applying the open water likelihood (OWL) algorithm on MODIS surface reflectance imagery. For the estimation of the associated water volumes, three different data driven methodologies were compared, all of which use digital elevation models (DEM) to obtain water depths. These involve the obtaining of the maximum elevation in the flooded polygons, and the use of the Cohen and Doble algorithms. Two DEM products were used, a 5 m resolution LiDAR dataset of the floodplain in the catchment and the 1 second SRTM derived elevation model. Flood volumes were compared with rainfall volumes and the discharge at several gauge stations located at different reaches of the river. Return periods were obtained from the probabilities of pixels being inundated in a year. The relation between flood volume estimations and the stream discharge varied depending on the gauge position in the catchment. Flood volume estimation was improved using methods that took into account the flood pattern connection with the channels. A single flood frequency curve was developed for the entire catchment.
Statement of Contribution of Co-Authors

This chapter has been written as a conference article. The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
5. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for this publication is:


<table>
<thead>
<tr>
<th>Contributors</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ignacio Fuentes</strong></td>
<td>Original idea; Coding; Data analysis; Writing</td>
</tr>
<tr>
<td><strong>Jose Padarian</strong></td>
<td>Edition; Suggestions</td>
</tr>
<tr>
<td><strong>R. Willem Vervoort</strong></td>
<td>Supervision; Edition; Suggestions</td>
</tr>
</tbody>
</table>
Chapter 5. Comparing volume estimates from flood detection

5.1. Introduction

Globally unsustainable production and exploitation rates create significant uncertainty about the future of the planet (Schewe et al., 2014). Quantifying water resources to assist water management plans or risk assessments therefore remains a priority (Poff et al., 2016).

Remote sensing use in estimating water resources has been increasing (Mueller et al., 2016), and can be applied together with direct measurements to assess the state and temporal variation of processes (Doble et al., 2016; Siev et al., 2016). The advantage of satellite imagery, relative to many other data sources is its ability to capture spatial variability of features at the surface (Mueller et al., 2016). While gauge and meteorological stations collect data at specific locations (Alsdorf et al., 2007), satellites cover a large spatial area (Gupta, 2018). On the other hand, satellites also measure the same spot at different times, which gives a time series (Siev et al., 2016). There are several satellites with different resolutions, capturing different wavelength ranges (Gupta, 2018; Schmugge et al., 2002). In addition, not all satellites have operated over the same timeframe, with some of the more detailed spatial products being relatively recent (Sentinel; Nagler et al. 2015).

While water detection from space captures the spatial extent of water on the surface, it does not necessarily provide water quantities in flood events (Zhou et al., 2017). As a result, data driven methods have been developed to estimate inundation volumes by coupling flood extension imagery with gauge station measurements, bathymetry or digital elevation models (DEM; Cohen et al., 2017; Doble et al., 2014; Siev et al., 2016). Several approaches have been tested, but the performance against field measurements has mostly been through spatial snapshots rather than studying the temporal dynamics (Cohen et al., 2017), or have been studied mostly in the short term (Frappart et al., 2005; Siev et al., 2016).

The tendency for short period flood studies, using remote sensing, is mainly caused by difficulties of acquiring large amounts of remote sensing information and related preprocessing (Ma et al., 2015). However, an important innovation has been the development of the Google Earth Engine platform, which contains multi-petabyte processed geospatial datasets that are being updated and uploaded constantly (Gorelick et al., 2017).

The second main difficulty to assess inundation volumes is the lack of a reference to compare the results (Oreskes et al., 1994). Most of the time, information from gauge stations, dams, or the inputs of water to the system, such as rainfall, are used, all of which allow a rough water availability estimation (Alsdorf et al., 2007). However, more alternatives have been developed in the last years, such as the use of gravity...
Chapter 5. Comparing volume estimates from flood detection

satellite imagery (GRACE), and in the next years the launch and operation of the Surface Water and Ocean Topography mission (SWOT) is expected to produce more accurate estimations of surface water in the future (Fu and Ubelmann, 2014).

The main objective of this study was to compare three different existing methodologies for water volume estimation with a water detection algorithm through Google Earth Engine to study long-term flood volume dynamics at the catchment scale.

5.2. Materials and methods

5.2.1. Study catchment

The selected 42,000 km² catchment is the Namoi, located in the north of NSW, Australia, formed by the Namoi river flowing westward. It has several major dams, storing water and regulating flow downstream. Another main hydrological characteristic are on-farm “ring tank” dams for irrigation, which can be up to 100 ha.

The yearly mean precipitation in the catchment is 800 mm, whilst the mean annual potential evapotranspiration is around 1300 mm (McCallum et al., 2010). Floods in the Namoi catchment are periodic natural events with a significative environmental role for local wetlands (Green et al., 2011), but also causing potential economical and humanitarian losses. The catchment elevation ranges from 125 m.a.s.l., in the west, and increases eastwards up to 1,501 m.a.s.l. This height gradient is one of the main factors that influences areas prone to inundation in the catchment.

5.2.2. Data selection

Data from several gauging stations in the catchment was used in this study. Daily stream levels and discharges are freely available at the waterinfo webpage, from the New South Wales government (http://waterinfo.nsw.gov.au/). Daily rainfall grids at a resolution of 0.05 degrees were obtained from the Bureau of Meteorology of the Australian Government (http://www.bom.gov.au). The grids were clipped at the extension of the basin and rainfall was daily summed to estimate the total volume of water entering the catchment. Satellite imagery used were the daily and the 8 days composite surface reflectance datasets obtained from the MODIS Terra satellite (MOD09GA and MOD09A1, respectively) from 2000 to 2018. These datasets, available in the Google Earth Engine platform, were preprocessed removing clouds,
Chapter 5. Comparing volume estimates from flood detection

shadows, and smoke from fires using the StateQA Bitmask band contained in the images, which contains a classification of different land features captured by the satellite sensors. The images were subsequently masked in order to remove pixels that were affecting the detection of water at the margin of tiles and in the range of the shortwave infrared wavelength. Additionally, permanent water bodies were masked using the Land Water Mask Derived from MODIS and SRTM (MOD44W). A SRTM Derived Hydrological Elevation Model at 1 arcsec resolution and a 5 m Digital Elevation Model (DEM) of Australia derived from LiDAR, both available in the Google Earth engine platform, were used to obtain the water depths in flooded areas, and to derive a multiresolution index of valley bottom flatness (MrVBF) map of the catchment. Both DEMs were used and compared. The LiDAR DEM does not comprise the entire catchment, but only an area that covers some of the floodplains. In spite of this, the vertical resolution of this DEM is significantly more accurate than that from the SRTM DEM, which does cover the entire extension of the catchment.

5.2.3. Inundation estimations

In order to estimate the flood extent, the Open Water Likelihood (OWL) algorithm (Guerschman et al. 2011) was used, which gives the probability of finding water in each pixel. The MrVBF used as input in the algorithm was derived from the SRTM DEM, and a threshold of 0.87 was applied to the OWL data to select inundated areas because this value gave the best estimates of actual flood extent (results not shown).

To estimate the water volumes from the flood extension data obtained, three methods that use DEMs or bathymetric images were utilised, allowing to get the depth of water at surface, which was subsequently multiplied by the inundation areas at each pixel. For all methods, the inundation area was constrained to the area covered by the LiDAR imagery in order to compare the results using both, the SRTM and the LiDAR DEMs. The water depth estimation methods used are described below:

- The first method, subsequently referred to as Maximum, assumes that the water surface during floods is flat (Siev et al., 2016). Polygons of inundated areas are obtained and overlain by the DEM, and the maximum elevation of water in the perimeter of polygons is assumed to be the surface water elevation. Then, the DEM is subtracted to get the water depth.
- In the second method, subsequently referred to as Doble, the DEM was converted into an array of 100 matrices by obtaining the percentiles of all DEM pixels (30 m resolution) included in each MODIS pixel (500 m resolution). Then, an elevation image was created by selecting the elevation percentile of the array in each pixel corresponding to the probability estimated by the OWL.
algorithm. The original DEM was subsequently subtracted from these
elevations to obtain the water depth at each pixel (Doble et al., 2014). All
negative values calculated were converted to 0, and only areas with an OWL
threshold of 0.87 were classified as inundated.

- The last method, subsequently referred to as Cohen, was developed by Cohen
  et al. (2017). It involves a conversion of inundated areas into polygons to
  obtain the elevations at the perimeter of polygons, and subsequently it applies
  a focal mean in a series of iterations to populate the area inside the polygons
  with water elevations. The final stage implies subtracting the water elevations
  from the original DEM to get the water depths. Negative water depths are
  converted to 0, and a final focal mean with a kernel of 3 pixels is carried out
  to smooth any abrupt change on the water elevations.

### 5.2.4. Occurrence probability and statistical analysis

Using the flooded areas from the OWL algorithm probability maps were calculated
out in Google Earth Engine. For each year, the maximum inundated area was
estimated. Then, the probability of inundated pixels was obtained using Kuczera and
Franks (2016; Eq. 5.1):

\[
p_{i,j} = \frac{m_{i,j}^{-0.4}}{n+0.2} \tag{5.1}
\]

where \(p_{i,j}\) is the occurrence probability for each pixel, being \(i\) and \(j\) the pixel positions
in the image, \(m\) corresponds to number of occurrences and \(n\) to the number of years
on record. The probability used differs from the classical Weibull equation in the
constants that it uses since it leads to unbiased estimations of quantiles for different
families of probability. Then, the return period is calculated as the inverse of the
probability of occurrences. From the occurrence probability maps an estimate of
inundated volumes was calculated, which gave flood frequency curves at a
catchment scale. These curves were compared with flood frequency curves
estimated from gauge stations using the same methodology.

### 5.3. Results and discussion

The algorithms result in different inundation volumes (Figure 5.1). In general, the
algorithm that takes into account the maximum water levels, predicts inundation
values that are around two orders of magnitude greater than the Cohen algorithm,
and at least one order of magnitude higher than using the OWL probabilities (Doble
et al, 2014).
Chapter 5. Comparing volume estimates from flood detection

Figure 5.1. Time series of inundation volumes estimated and daily discharges at Walgett.

All the inundated volume time series capture the peak of discharges in the stream, and also present a seasonal behaviour. The logarithmic scale allows to detect not just flood events, but also some periods where there is limited surface water, probably associated to drought periods, which can be observed in 2003, 2007 and 2015. In order to analyze how inundation volumes respond to rainfall, a cross correlation was carried out (Figure 5.2).

Figure 5.2. Cross correlation between inundated volumes and daily rainfall.
As it is expected, there is a delay in the response of floods to rain, with the highest correlations found after 10 days. In this case, the Cohen and Doble algorithms have very similar correlations, which are in general higher than those obtained using the maximum DEM values.

Taken into account the lagged response observed, a scatter plot between monthly accumulated rainfall and monthly accumulated inundated volumes can be assessed for each methodology (Figure 5.3). Accumulated volumes should be expected to be always lower than accumulated rainfall, considering that rain corresponds to the input of water to the system and that surface waters are just a fraction of the available water of the system.

![Figure 5.3. Scatterplots of monthly accumulated rainfall and monthly accumulated inundation volumes.](image)

In the case of the maximum elevation methodology used, some accumulated inundation volumes are significantly higher than accumulated rainfall. This was also tested by assessing the maximum and minimum values of the perimeter of inundation polygons. In most cases, and for small inundation polygons, the difference between the minimum and maximum elevation values was not significant (lower than 0.5 m). However, during flood events, big inundation polygons surrounding the main stream could be observed, and the perimeter had elevations ranging in some cases higher than 10 m, which means that this methodology fails to meet the original assumption of flat water surfaces, and implying that part of the inundation volumes detected is still flowing through the main streams (Frappart et al., 2005).
The relationship between the inundated volumes and the daily discharges in the stream gauges of the Namoi river at Bugilbone and Walgett (both in the lower western part of the catchment, Figure 5.4) indicates a clear relationship with correlation coefficients ranging from 0.57 to 0.93, depending on the methodology. The lowest correlation is for the maximum elevation methodology, whilst the other two methods have quite similar correlation coefficients.

**Figure 5.4.** Scatterplots of inundated volumes against daily discharge volumes in the Namoi river at Bugilbone (above) and Walgett gauge stations (below).

A cross correlation between daily discharge at all stations in the catchment and daily flood volumes in the LiDAR DEM locations identified the maximum cross correlation coefficients (left below, Figure 5.5) and associated time lag between the flood volume and occurrence of discharge (left above, Figure 5.5). Is it clear that in the lower (western) part of the catchment, lags are smaller and even positive (indicating the flooding occurs before the gauge registers an increase in flow). These areas which receive all the upstream flow, also have higher correlation values than the gauges located in the east of the catchment (higher elevations). The eastern (upstream) part of the catchment, also indicates stronger negative lags, meaning the flooding follows the increase in flow at the gauge.
Different locations in the catchment have different lags between rainfall and flooding, which can assist with forecasting, but also gives information about the rate of surface water movement in the catchment (Alsdorf et al., 2007).

Figure 5.5. Maximum cross correlation coefficients between daily discharge at different locations in the catchment and inundation volumes (left below) obtained within different time lags (left above), and occurrence probability maps using daily (MOD09GA; right above) and 8 days composite (MOD09A1; right below) imagery.

The deviation of linear regression slopes, obtained between daily discharge and inundation volumes with the different time lags used above, from a 1:1 line can be used to estimate the proportion of water losses other than stream discharge (evapotranspiration and recharge). This varies by reach of the river, and the distance to the flood sinks (Figure 5.6).
Chapter 5. Comparing volume estimates from flood detection

Figure 5.6. Proportion of water losses other than stream discharge obtained through the deviation of regression slopes from a 1:1 line.

Thus, whilst in the main reach of the Namoi river from Boggabri downstream a fairly even proportion of water losses occurs, Pian Creek and elevated sectors of the catchment have a much higher proportion of losses by recharge and evapotranspiration, with some locations near Wee Waa showing very high water loss. In this area, it has been identified a significant surface-groundwater interaction (Kelly et al., 2009), which is recharging the Namoi aquifer, and would shift the slope of the regressions. Negative fractions are observed at two gauges, which also have the lowest correlation coefficients between flood and discharges in the catchment. These are located immediately downstream the Chaffey dam, one of the major water reservoirs in the basin, which also regulates the flow downstream (Green et al., 2011).

Occurrence probability images obtained from the OWL algorithm by using daily and composite imagery (Figure 5.5) were processed to estimate the frequency of flood events in the catchment and compared with flood frequency curves obtained from the Namoi river at Walgett and Bugilbone gauging stations (Kuczera and Franks, 2016; Figure 3.7). Flood frequency curves from the Cohen and the Doble algorithms are significantly different, being much lower for the Cohen methodology, especially at small return periods. Additionally, whilst in the Cohen methodology the use of the LiDAR DEM always leads to smaller inundation volumes, in the Doble methodology the flood frequency curves overlap, and the difference between the curves is smaller. The inundation volumes for the entire catchment (Cohen basin and Doble basin in Figure 5.7a) are significantly higher than those obtained for the extension covered by the LiDAR images, but follow the same general trend. The flood frequency curves from the river at Walgett and Bugilbone indicate similar behaviour (Figure 5.7b).
Figure 5.7. Flood frequency curves obtained from the temporal analysis of flood volume estimations (left), and from daily discharges in the Namoi river at Bugilbone and at Walgett (right).

5.4. Conclusions

A surface water detection algorithm and three water depth methodologies were coupled in order to get time series of inundation volume estimations at a catchment scale. The estimated volumes for all methods presented seasonal variation and captured peak discharge events. However, the method that assumes the maximum elevation of flooded polygons proved to be the worst as the water elevation of them demonstrated not to meet the assumption of a flat surface, and therefore overpredicts inundated volumes at big flood events. The other methodologies, despite differences in volume estimates, did not show significant differences in temporal behaviour. Rainfall is related to flooded volumes, but this relationship is higher several days after rainfall events. The relationship between inundation volumes and discharges at different locations in the catchment varies depending on the location. In general, downstream stations presented higher correlations and also a delay in the flow response to floods, whilst upstream stations presented lower correlation coefficients and its discharge peaks precedes the flood occurrence. In addition, the slope of the linear regression between daily discharges and daily inundation volumes and its deviation from a 1:1 line can allow to obtain the proportion of other water losses, such as recharge or evapotranspiration. This changes according to the position in the catchment, and especially depending on the reach where the stations are located, meaning that the distance to flood sinks and hydraulic properties of the sediments are affecting the flood-discharge response. From water occurrence probability maps, flood frequency curves were obtained for the entire catchment, presenting a behavior similar to curves calculated from gauge stations. The results obtained can help to the understanding of the hydrological cycle.
and the processes related, and also to define management plans and hydraulic design thresholds.

5.5. References


Chapter 5. Comparing volume estimates from flood detection


Chapter 6

Word embeddings for application in geosciences: development, evaluation and examples of soil-related concepts

Abstract
A large amount of descriptive information is available in geosciences. This information is usually considered subjective and ill-favoured compared with its numerical counterpart. Considering the advances in natural language processing and machine learning, it is possible to utilise descriptive information and encode it as dense vectors. These word embeddings, which encode information about a word and its linguistic relationships with other words, lay on a multidimensional space where angles and distances have a linguistic interpretation. We used 280,764 full-text scientific articles related to geosciences to train a domain-specific language model capable of generating such embeddings. To evaluate the quality of the numerical representations, we performed three intrinsic evaluations: the capacity to generate analogies, term relatedness compared with the opinion of a human subject, and categorisation of different groups of words. As this is the first attempt to evaluate word embedding for tasks in the geosciences domain, we created a test suite specific for geosciences. We compared our results with general domain embeddings commonly used in other disciplines. As expected, our domain-specific embeddings (GeoVec) outperformed general domain embeddings in all tasks, with an overall performance improvement of 107.9%. We also presented an example where we successfully emulated part of a taxonomic analysis of soil profiles that was originally applied to soil numerical data, which would not be possible without the use of embeddings. The resulting embedding and test suite will be made available for other researchers to use and expand upon.
Statement of Contribution of Co-Authors

This chapter has been written as a journal article. The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
5. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for this publication is:


<table>
<thead>
<tr>
<th>Contributors</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>José Padarian</td>
<td>Conceptualization; Data analysis; Writing</td>
</tr>
<tr>
<td>Ignacio Fuentes</td>
<td>Data analysis; Writing</td>
</tr>
</tbody>
</table>
6.1. Introduction

Machine learning (ML) methods have been used in many fields of geosciences (Lary et al., 2016) to perform tasks such as the classification of satellite imagery (Maxwell et al., 2018), soil mapping (McBratney et al., 2003), mineral prospecting (Caté et al., 2017), and flood prediction (Mosavi et al., 2018). Owing to their capability to deal with complex non-linearities present in the data, ML usually outperforms more traditional methods in terms of predictive power. The application of ML in geosciences commonly prioritises numerical or categorical data over qualitative descriptions, which are usually considered subjective in nature (McBratney and Odeh, 1997). However, the resources that have been invested in collecting large amounts of descriptive information from pedological, geological, and other fields of geosciences must be taken into account. Neglecting descriptive data due to their inconsistency seems wasteful; however, natural language processing (NLP) techniques, which involve the manipulation and analysis of language (Jain et al., 2018), have rarely been applied in geosciences.

For soil sciences, NLP opens the possibility to use a broad range of new analyses. Some examples include general, discipline-wide methods such as automated content analysis (Nunez-Mir et al., 2016) or recommendation systems (Wang and Blei, 2011) which can take advantage of the current literature. More specific cases could take advantage of big archives of descriptive data, such as those reported by Arrouays et al. (2017). The authors mention examples such as the Netherlands with more than 327,000 auger descriptions covering agricultural, forest, and natural lands, or the north-central US with 47,364 pedon descriptions covering eight states.

Approaches to deal with descriptive data include the work of Fonseca et al. (2002), who proposed the use of ontologies to integrate different kinds of geographic information. At the University of Colorado, Chris Jenkins created a structured vocabulary for geomaterials (https://instaar.colorado.edu/~jenkinsc/dbseabed/resources/geomaterials/, last access: 12 July 2019) using lexical extraction (Miller, 1995), names decomposition (Peckham, 2014), and distributional semantics (Baroni et al., 2012) in order to characterise word terms for use in NLP and other applications. A different approach, perhaps closer to the preferred quantitative methods, is the use of dense word embeddings (vectors) which encode information about a word and its linguistic relationships with other words, positioning it on a multidimensional space. The latter is the focus of this study.

There are many general-purpose word embeddings trained on large corpora from social media or knowledge organisation archives such as Wikipedia (Pennington
et al., 2014; Bojanowski et al., 2016). These embeddings have been proven to be useful in many tasks such as machine translation (Mikolov et al., 2013a), video description (Venugopalan et al., 2016), document summarisation (Goldstein et al., 2000), and spell checking (Pande, 2017). However, for field-specific tasks, many researchers agree that word embeddings trained on specialised corpora can more successfully capture the semantics of terms than those trained on general corpora (Jiang et al., 2015; Pakhomov et al., 2016; Roy et al., 2017; Nooralahzadeh et al., 2018; Wang et al., 2018).

As far as we are aware, this is the first attempt to develop and evaluate word embedding for the geosciences domain. This paper is structured as follows: first, we define what word embeddings are, explaining how they work and showing examples to help the reader understand some of their properties; second, we describe the text data used and the pre-processes required to train a language model and generate these word embeddings (GeoVec); third, we illustrate how a natural language model can be quantitatively evaluated and we present the first test dataset for the evaluation of word embeddings specifically developed for the geosciences domain; fourth, we present results of an intrinsic evaluation of our language model using our test dataset and we explore some of the characteristics of the multidimensional space and the linguistic relationships captured by the model using examples of soil-related concepts; and finally, we present a simple, illustrative example of how the embedding can be used in a downstream task.

### 6.2. Word embeddings

Word embeddings have been commonly used in many scientific disciplines, thanks to their application in statistics. For example, one-hot encodings (Figure 6.1), also known as “dummy variables”, have been used in regression analysis since at least 1957 (Suits, 1957). In one-hot encoding, each word is represented by a vector of length equal to the number of classes or words, where each dimension represents a feature. The problem with this representation is that the resulting array is sparse (mostly zeros) and very large when using large corpora; in addition, it also presents the problem of poor estimation of the parameters of the less-common words (Turian et al., 2010). A solution for these problems is the use of unsupervised learning to induce dense, low-dimensional embeddings (Bengio, 2008). The resulting embeddings lie on a multidimensional space where angles and distances have a linguistic interpretation.
These dense, real vectors allow models, especially neural networks, to generalise to new combinations of features beyond those seen during training due to the properties of the vector space where semantically related words are usually close to each other (LeCun et al., 2015). As the generated vector space also has properties such as addition and subtraction, Mikolov et al. (2013b) gives some examples of calculations that can be performed using word embedding. For instance the operation $\text{vec("Berlin")} - \text{vec("Germany")} + \text{vec("France")}$ generates a new vector. When they calculated the distance from that resulting vector to all the words from the model vocabulary, the closest word was “Paris”. Figure 6.2 presents a principal component analysis (PCA) projection of pairs of words with the country–capital relationship. Without explicitly enforcing this relationship when creating the language model, the resulting word embeddings encode the country–capital relationship due to the high co-occurrence of the terms. In Figure 6.2 it is also possible to observe a second relationship, geographic location, where South American countries are positioned to the right, European countries in the middle, and (Eur-) Asian countries to the left.
6.3. Data, text pre-processing, and model training

6.3.1. Corpus

The corpus was generated by retrieving and processing 280,764 full-text articles related to geosciences. We used the Elsevier ScienceDirect APIs (application
programming interfaces) to search for literature that matched the terms listed in Table 1, which cover a broad range of topics. These terms were selected based on their general relationship with geosciences and specifically soil science. We also included Wikipedia articles that list and concisely define some concepts such as types of rocks, minerals, and soils, providing more context than a scientific publication, considering that the model depends on words co-occurrences. We downloaded the text from Wikipedia articles “List_of_rock_types”, “List_of_minerals”, “List_of_landforms”, “Rock_(geology)”, “USDA_soil_taxonomy”, and “FAO_soil_classification”, and also all of the Wikipedia articles linked from those pages.

**Table 6.1.** Search terms used to retrieve full-text articles from Elsevier ScienceDirect APIs.

<table>
<thead>
<tr>
<th>Search terms</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acrisol</td>
<td>Geosciences</td>
<td>Permafrost</td>
</tr>
<tr>
<td>Alfisol</td>
<td>Groundwater</td>
<td>Petrology</td>
</tr>
<tr>
<td>Allophane</td>
<td>Gypsisols</td>
<td>Podzols</td>
</tr>
<tr>
<td>Andisol</td>
<td>Histosol</td>
<td>Sedimentary</td>
</tr>
<tr>
<td>Andosols</td>
<td>Hydrogeology</td>
<td>Sedimentary mineralogy</td>
</tr>
<tr>
<td>Aridisol</td>
<td>Igneous petrology</td>
<td>Sedimentary petrology</td>
</tr>
<tr>
<td>Chernozems</td>
<td>Imogolite</td>
<td>Sedimentary rocks</td>
</tr>
<tr>
<td>Entisol</td>
<td>Inceptisol</td>
<td>Sedimentology</td>
</tr>
<tr>
<td>Environmental geology</td>
<td>Lithology</td>
<td>Soil classification</td>
</tr>
<tr>
<td>Field geology</td>
<td>Metamorphic petrology</td>
<td>Spodosol</td>
</tr>
<tr>
<td>Gelisol</td>
<td>Mineralogy</td>
<td>Stratigraphy</td>
</tr>
<tr>
<td>Geochemistry</td>
<td>Mollisol</td>
<td>Ultisol</td>
</tr>
<tr>
<td>Geology</td>
<td>Oxisol</td>
<td>Vertisol</td>
</tr>
<tr>
<td>Geomaterials</td>
<td>Peatland</td>
<td>Volcanic soil</td>
</tr>
<tr>
<td>Geomorphology</td>
<td>Pedogenesis</td>
<td></td>
</tr>
<tr>
<td>Geophysics</td>
<td>Pedology</td>
<td></td>
</tr>
</tbody>
</table>

**6.3.2. Pre-processing**

The corpus was split into sentences which were then pre-processed using a sequence of commonly used procedures including the following:

- removing punctuation,
- lowercasing,
removing digits and symbols, and

• removing (easily identifiable) references.

The cleaned sentences were then tokenised (split into words). In order to decrease the complexity of the vocabulary, we lemmatised all nouns to their singular form and removed all the words with less than three characters. We also removed common English words such as “the”, “an”, and “most” as they are not discriminating and unnecessarily increase the model size and processing time (a full list of the removed “stop words” can be found in the documentation of the NLTK Python library; Bird and Loper, 2004). Finally, we excluded sentences with less than three words. The final corpus has a vocabulary size of 701,415 (unique) words and 305,290,867 tokens.

### 6.3.3. Model training

For this work, we used the GloVe (Global Vectors) model (Pennington et al., 2014), developed by the Stanford University NLP group, which achieved great accuracy on word analogy tasks and outperformed other word embedding models on similarity and entity recognition tasks. As with many NLP methods, GloVe relies on ratios of word–word co-occurrence probabilities in the corpus. To calculate the co-occurrence probabilities, GloVe uses a local context window, where a pair of words $d$ words apart contributes $\frac{1}{d}$ to the total count. After the co-occurrence matrix $X$ is calculated, GloVe minimises the least-squares problem:

$$\sum_{i,j=1}^{V} f(X_{ij})(w_i^T \hat{w}_j + b_i + \hat{b}_j - \log X_{ij})^2$$ (6.1)

where $X_{ij}$ is the co-occurrence between the target words $i$ and the context word $j$, $V$ is the vocabulary size, $w_i$ is the word embedding, $\hat{w}_j$ is a context word embedding, $b_i$ and $\hat{b}_j$ are biases for $w_i$ and $\hat{w}_j$, respectively, and $f(X_{ij})$ is the weighting function:

$$f(X_{ij}) = \begin{cases} (X_{ij}/x_{\max})^\alpha & \text{if } X_{ij} < x_{\max} \\ 1 & \text{otherwise} \end{cases}$$ (6.2)

that assures that rare and frequent co-occurrences are not overweighted. Pennington et al. (2014) recommend using the values 0.75 for the smoothing parameter $\alpha$ and 100 for the maximum cutoff count $x_{\max}$. 

155
We trained the model during 60 epochs, where 1 epoch corresponds to a complete pass through the training dataset. During the training phase, we experimented using embedding with different number of components (dimensions) and different context window sizes. Here we present the results for 300 components and a context window of size 10, which represents a good balance between model size, training time and performance.

### 6.3.4. Evaluating of word embeddings

Given the characteristic of the vector space, the most common method to evaluate word embeddings is to assess their performance in tasks that test if semantic and syntactic rules are properly encoded. Many studies have presented datasets to perform this task. Rubenstein and Goodenough (1965) presented a set of 65 noun synonyms to test the relationship between the semantic similarity existing between a pair of words and the degree to which their contexts are similar. More recent and larger test datasets and task types have been proposed (Finkelstein et al., 2002; Mikolov et al., 2013c; Baroni et al., 2014), but they have all been designed to test general domain vectors. Because this work aims to generate embeddings for the geosciences domain, we developed a test suite to evaluate their intrinsic quality in different tasks, which are described below.

**Analogy:**

Given two related pairs of words, $a:b$ and $x:y$, the aim of the task is to answer the question “$a$ is to $x$ as $b$ is to?” The set includes 50 quartets of words with different levels of complexity, from simple semantic relationships to more advance syntactic relations. In practice, it is possible to find $y$ by calculating the cosine similarity between the differences of the paired vectors:

$$
\frac{(v_b - v_a) \cdot (v_y - v_x)}{||v_b - v_a|| \cdot ||v_y - v_x||}
$$

(6.3)

In this case, $v_y$ is the embedding for each word of the vocabulary and $y$ is the word with the highest cosine similarity. Some examples of analogies are “moraine is to glacial as terrace is to []? (fluvial)”, “limestone is to sedimentary as tuff is to []? (volcanic)”, and “chalcanthite is to blue as malachite is to []? (green)“.

We estimated the top-1, top-3, top-5, and top-10 accuracy score, recording a positive result if $y$ was within the first 1, 3, 5, or 10 words returned by the model, respectively.
Chapter 6. Word embeddings for applications in geosciences

**Relatedness:**

For a given pair of words \((a, b)\), a score of zero or one is assigned by a human subject if the words are unrelated or related, respectively. The set includes 100 pairs of scored pairs of words. The scores are expected to have a high correlation with the cosine similarity between the embeddings of each pair of words. In this work, we used the Pearson correlation coefficient to evaluate the model against annotations made by three people with a geosciences background.

**Categorisation:**

Given two sets of words \(S_1 = \{a, b, c, \ldots\}\) and \(S_2 = \{x, y, z, \ldots\}\), this test should be able to correctly assign each word to its corresponding group using a clustering algorithm. We provide 30 tests with 2 clusters each. We estimated the \(\nu\)-measure score (Rosenberg and Hirschberg, 2007), which takes the homogeneity and completeness of the clusters into account, after projecting the multidimensional vector space to a two-dimensional PCA space and performing a \(k\)-means clustering. Given that \(k\)-means is not deterministic (when using random centroids initiation), we used the mean \(\nu\)-measure score of 50 realisations.

We compared our results with general domain vectors trained on Wikipedia articles (until 2014) and the Gigaword v5 catalogue, which comprises 6 billion tokens and is provided by the authors of GloVe at [https://nlp.stanford.edu/projects/glove/](https://nlp.stanford.edu/projects/glove/) (last access: 12 July 2019).

**6.3.5. Illustrative example**

In order to illustrate the use of word embedding in a downstream application, we decided to emulate part of the analysis of a soil taxonomic system performed by Hughes et al. (2017). They used 23 soil variables (e.g. sand content and bulk density), where the majority were numerical and continuous except for two binary variables representing the presence or absence of water or ice. Those variables correspond to the representation of horizons from soil profiles, which were then aggregated (mean) at different taxonomic levels to obtain class centroids.

Our analysis was similar, but, instead of using soil variables, we used the word embedding corresponding to the textual description of 10 000 soil profile descriptions downloaded from the United States Department of Agriculture–Natural Resources Conservation Service (USDA-NRCS) web site for official soil series descriptions and series classification. The descriptions were pre-processed utilising the same pipeline used for the corpus (Sect. 6.3.2). After obtaining the embeddings for each token in the descriptions, we calculated the mean values per profile, which can be considered as an embedding at the profile level. The profiles and their corresponding 300-
dimensional embeddings were aggregated at the great group (GG) level (soil taxonomy) and a mean embedding value was estimated (equivalent to the centroids obtained by Hughes et al., 2017). After projecting the GG embeddings into a two-dimensional PCA space, we computed the convex-hull per soil order (smaller convex polygon needed to contain all the GG points for a particular soil order) as a way of visualising their extent.

6.4. Results and discussion

6.4.1. Co-occurrence

Before training the language model, the first output of the process is a co-occurrence matrix. This matrix encodes useful information about the underlying corpus (Heimerl and Gleicher, 2018). Figure 6.3 shows the co-occurrence probabilities of soil taxonomic orders and some selected words. It is possible to observe that concepts generally associated with a specific order co-occur in the corpus, such as the fact that soil cracks are features usually present in Vertisols, or that Andisols are closely related to areas with volcanic activity.

![Co-occurrence probability matrix of soil orders (USDA) and selected words.](image)

**Figure 6.3.** Co-occurrence probability matrix of soil orders (USDA) and selected words.

This information can also be used to guide the process of generating a domain-specific model. In our case, in an early stage of this study, the terms “permafrost” and “Gelisol” presented a very low co-occurrence probability, which was a clear sign of the limited topic coverage of the articles at that point.
6.4.2. Intrinsic evaluation

The results of the intrinsic evaluation indicate that our domain-specific embeddings (GeoVec) performed better than the general domain embeddings in all tasks (Table 2), increasing the overall performance by 107.9%. This is an expected outcome considering the specificity of the tasks. For the analogies, we decided to present the top-1, 3, 5, and 10 accuracy scores because, even if the most desirable result is to have the expected word as the first output from the model, in many cases the first few words are closely related or they are synonyms. For instance, for the analogy “fan is to fluvial as estuary is to []? (coastal)”, the first four alternatives are “tidal”, “river”, “estuarine”, and “coastal”, which are all related to an estuary.

In the relatedness task, the three human annotators had a high inter-annotator agreement (multi-kappa = 98.66%; as per Davies and Fleiss, 1982), which was expected as the relations are not complex for someone with a background in geosciences. As we keep working on this topic, we plan to extend the test suite with more subtle relations.

Table 6.2. Evaluation scores for each task for our domain-specific (GeoVec) and general domain embeddings (Stanford). For the analogy task, top-1, 3, 5, and 10 represents the accuracy if the expected word was within the first 1, 3, 5, or 10 words returned by the model. For the relatedness task, the score represents the absolute value of the Pearson correlation (mean of the three human subjects). For the categorisation task, the score represents the mean value of 50 v-measure scores. The possible range of all scores is zero to one, where higher is better.

<table>
<thead>
<tr>
<th>Task</th>
<th>GeoVec</th>
<th>Stanford</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analogy (top-1)</td>
<td>0.39</td>
<td>0.22</td>
</tr>
<tr>
<td>Analogy (top-3)</td>
<td>0.78</td>
<td>0.37</td>
</tr>
<tr>
<td>Analogy (top-5)</td>
<td>0.90</td>
<td>0.41</td>
</tr>
<tr>
<td>Analogy (top-10)</td>
<td>0.92</td>
<td>0.49</td>
</tr>
<tr>
<td>Relatedness</td>
<td>0.61</td>
<td>0.23</td>
</tr>
<tr>
<td>Categorisation</td>
<td>0.75</td>
<td>0.38</td>
</tr>
<tr>
<td>Overall</td>
<td>0.73</td>
<td>0.35</td>
</tr>
</tbody>
</table>

It was possible to observe an increase in the overall performance of the embeddings (calculated as the mean of the analogy – top-5, relatedness, and categorisation tasks) as we added more articles, almost stabilising around 300 million tokens, especially for the analogy task (Figure 6.4). For domain-specific embeddings, this limit most likely varies depending on the task and domain. For instance, Pedersen
et al. (2007), measuring semantic similarity and relatedness in the biomedical domain, found a limit of around 66 million tokens.

![Figure 6.4](image)

**Figure 6.4.** Overall performance of the embeddings versus number of tokens used to construct the co-occurrence matrix. The improvement limit is around 300 million tokens. For future comparisons, this limit corresponds to approximately 280 000 articles, 22.5 million sentences and 700 000 unique tokens.

The improvement over the general domain embeddings has also been reported in other studies. Wang et al. (2018) concluded that word embeddings trained on biomedical corpora can more suitably capture the semantics of medical terms than the embeddings of a general domain GloVe model. Also in a biomedical application, Jiang et al. (2015) and Pakhomov et al. (2016) reported similar conclusions. In the following sections, we explore the characteristics of the obtained embeddings, showing some graphical examples of selected evaluation tasks.

### 6.4.3. Analogy

A different way of evaluating analogies is to plot the different pairs of words in a two-dimensional PCA projection. Figure 6.5 shows different pairs of words which can be seen as group analogies. From the plot, any pair of related words can be expressed as an analogy. For example, from Figure 6.5a, it is possible to generate the analogy “claystone is to clay as sandstone is to [ ]? (sand)”, and the first model output is indeed “sand”.
As we showed in Figure 6.2, the embeddings encode different relationships with different degrees of sophistication. In Figure 6.5a it is possible to observe simple analogies, mostly syntactic as “claystone” contains the word “clay”. Figure 6.5b presents a more advanced relationship, where rock names are assigned to their corresponding rock type.

![Figure 6.5](image)

**Figure 6.5.** Two-dimensional PCA projection of selected words. Simple syntactic relationship between particle fraction sizes and rocks (a) and advanced semantic relationship between rocks and rock types (b).

### 6.4.4. Categorization

Similar to the analogies, the categorisation task can also present different degrees of complexity of the representations. In Figure 6.6a, k-means clustering can distinguish the two expected clusters of concepts, WRB (FAO, 1988) and soil taxonomy (USDA, 2010) soil classification names. Andisols and Andosols are correctly assigned to their corresponding groups but are apart from the rest, probably due to their unique characteristics. Vertisols are correctly placed in between the two groups as both have a soil type with that name. A second level of aggregation can be observed in Figure 6.6b. The k-means clustering correctly assigned the same soil groups from Fig. 6.6a into a general “soil types” group, different from “rocks”.

6.4.5. Other embedding properties

Interpolation of embeddings is an interesting exercise that allows to further explore if the corpus is well represented by the vector space. Interpolation has been used to generate a gradient between faces (Yeh et al., 2016; Upchurch et al., 2017), assist drawing (Baxter and ichi Anjyo, 2006), and transform speech (Hsu et al., 2017). Interpolation between text embeddings is less common. Bowman et al. (2015) analysed the latent vector space of sentences and found that their model was able to generate coherent and diverse sentences when sampling between two embeddings. Duong et al. (2016) interpolated between embeddings from two vector spaces trained on corpora from different languages to create a single cross-lingual vector space. The vector space from our model also presents similar characteristics.

In order to generate the interpolated embeddings, we obtained linear combinations of two-word embeddings using the formula:

\[ v_{int} = \alpha \cdot v_a + (1 - \alpha) \cdot v_b \]  

(6.4)

where \( v_{int} \) is the interpolated embedding, and \( v_a \) and \( v_b \) are the embeddings of the two selected words. By varying the value of \( \alpha \) in the range \([0,1]\), we generated a gradient of embeddings. For each intermediate embedding obtained by interpolation, we calculated the cosine similarity (Eq. 6.3) against all of the words in the corpus and selected the closest one.
The results showed coherent concepts along the gradients (Figure 6.7). The interpolation between “clay” and “boulder”, with fine and coarse size, respectively, yields a gradient of sizes as follows: clay < silt < sand < gravel < cobble < boulder. Another interpolation example, along with another type of relationship, is shown in Figure 6.7b. The interpolation between the rocks “slate” and “migmatite” yields a gradient of rocks with different grades of metamorphism as follows: slate < phyllite < schist < gneiss < migmatite.

![Figure 6.7](image)

Figure 6.7. Interpolated embedding in a two-dimensional PCA projection showing a size gradient (a) with clay < silt < sand < gravel < cobble < boulder; and a gradient of the metamorphism grade (b) with slate < phyllite < schist < gneiss < migmatite. Red and blue dots represent selected words (“clay” and “boulder”, and “slate” and “migmatite”) and black dots represent the closest word (cosine similarity) to the interpolated embeddings.

### 6.4.6. Illustrative example

As a final, external evaluation of the embedding, we estimated average embeddings for each great group (soil taxonomy) of soils from 10,000 soil profiles descriptions. The convex-hulls at the soil order level (Figure 6.8) show the same pattern reported by Hughes et al. (2017). Thanks to the unique characteristic of Histosols and the high diversity of this taxonomic group, they are easily differentiated in the two-dimensional projection, showing the highest variability. The rest of the soil orders overlap heavily as their differences are hard to simplify into a two-dimensional space. This overlap does not imply that the orders are not separable in a higher-dimensional space. Here we plot the first two principal components (PCs), which only account for 28.8% of the total variance. This is probably the same reason for the
overlap in the study by Hughes et al. (2017), as they account for 95% of the total variance only after 36 PCs (i.e. their plot, also using the first 2 PCs, probably explains a low proportion of the total variance, similar to our example).

![Convex-hulls of great group embeddings at the order level (soil taxonomy). Great group embeddings were obtained after averaging the embeddings of all the words in the descriptions of the profiles belonging to each great group. The convex-hulls were estimated from the two first principal components of the great group embeddings.](image)

**Figure 6. 8.** Convex-hulls of great group embeddings at the order level (soil taxonomy). Great group embeddings were obtained after averaging the embeddings of all the words in the descriptions of the profiles belonging to each great group. The convex-hulls were estimated from the two first principal components of the great group embeddings.

This example shows how, by using descriptions encoded as word embeddings, we were able to use the same methods as Hughes et al. (2017). In this case, if no soil variables (laboratory data) were available, word embeddings could be used instead. Ideally, we would expect to use word embeddings to complement numerical data and utilise valuable information included in the descriptive data. This is also possible with other approaches. Hughes et al. (2017) manually generated binary embeddings for the presence of ice and water. Another alternative to create embeddings is fuzzy logic. For example, McBratney and Odeh (1997) fuzzified categorical information from soil profiles such as depth, generating an encoding that represents the probability of belonging to different depth classes (e.g. a “fairly deep” soil could lay between the “shallow” and “deep” classes, with a membership of 0.5 to each class). The advantage of using word embeddings is that they are high-dimensional vectors that encode much more information applicable to many tasks, which would be difficult to replicate by manual encoding.
6.4.7. What do these embeddings actually represent?

It is worth discussing if word embeddings tell us anything about nature or if they really just tell us about the humanly constructed way that science is done and reported. A language model extracts information from the corpora to generate a representation in a high-dimensional space. This continuous vector space shows interesting features that relate words to each other, which were tested in multiple tasks designed to evaluate the syntactic regularities encoded in the embeddings. Considering the position that science is a model of nature (Gilbert, 1991) and assuming that the way we do and report science is a good representation of it, if the language model is a good representation of the corpora of publications, perhaps the derived syllogism – the language model is a good representation of nature – can be considered as true. Of course, the representation of a representation carries many impressions, but it is worth exploring its validity.

As shown by the linear combinations of embeddings (Figure 6.7), some aspects related to “size” are captured by the embeddings and, even if size categories are a human construct, they describe a measurable natural property. A more complex case is the illustrative example, where the embeddings capture some aspects of nature which are also captured by the numerical representation of its properties (in this case soil properties such as clay content, pH, among others). Given the results of the intrinsic evaluation of this work and others referenced throughout this article, it is probably impossible to generate the “perfect embeddings”. Even if we were able to process all of the written information available, and ignore the limitations of any language model, the embeddings would still be limited by our capacity to understand non-linear relationships (Doherty and Balzer, 1988) and, in turn, to understand nature.

Whether word embedding can give new insights about geosciences is still to be tested. Studies in other fields have shown some potentially new information. For instance, Kartchner et al. (2017) generated embeddings from medical diagnosis data and, after performing a clustering, they found clear links between some diagnoses related to advanced chronic kidney disease. Some of the relations are already known and accepted by the medical community, whereas others are new and are just starting to be studied and reported.

6.4.8. Future work

In the future, we expect to evaluate the effect of using our embeddings in more downstream applications (extrinsic evaluation). It is expected that domain-specific
Chapter 6. Word embeddings for applications in geosciences

embedding will necessarily improve the results of downstream tasks but this is not always the case. Schnabel et al. (2015) suggested that extrinsic evaluation should not be used as a proxy for a general notion of embedding quality, as different tasks favour different embeddings, but they are useful in characterising the relative strengths of different models. We also expect to expand the test suite with more diverse and complex tests, opening the process to the scientific community. Another interesting opportunity is the inclusion of word embeddings in numerical classification systems (Bidwell and Hole, 1964; Crommelin and De Gruijter, 1973; Sneath et al., 1973; Webster et al., 1977; Hughes et al., 2014) which try to remove subjectivity by classifying an entity (soil, rock, etc.) based on numerical attributes that describe its composition.

6.5. Conclusions

In this work we introduced the use of domain-specific word embeddings for geosciences (GeoVec), and specifically soil science, as a way to (a) reduce inconsistencies of descriptive data, and (b) open the alternative to include such data into numerical data analysis. Comparing the result with general domain embeddings, trained on corpus such as Wikipedia, the domain-specific embeddings performed better in common natural language processing tasks such as analogies, terms relatedness, and categorisation, improving the overall accuracy by 107.9%.

We also presented a test suite, specifically designed for geosciences, to evaluate embedding intrinsic performance. This evaluation is necessary to test if syntactic or semantic relationships between words are captured by the embeddings. The test suite comprises tests for three tasks usually described in the literature (analogy, relatedness, and categorisation) with different levels of complexity. As creating a set of gold standard tests is not a trivial task, we consider this test suite a first approach. In the future, we expect to expand the test suite with more diverse and complex tests and to open the process to the scientific community to cover different subfields of geosciences.

We demonstrated that the high-dimensional space generated by the language model encodes different types of relationships, using examples of soil-related concepts. These relationships can be used in novel downstream applications usually reserved for numerical data. One of these potential applications is the inclusion of embeddings in numerical classification. We presented an example were we successfully emulated part of a taxonomic analysis of soil profiles which was originally applied to soil numerical data. By encoding soil descriptions as word embeddings we were able to utilise the same methods used in the original application and obtain similar results. Ideally, we would expect to use word embeddings when no numerical data are
available or to complement numerical data to include valuable information included in the descriptive data.

6.6. References


Chapter 6. Word embeddings for applications in geosciences


Chapter 6. Word embeddings for applications in geosciences


Mikolov, T., tau Yih, W., Zweig, G., 2013c. Linguistic regularities in continuous space word representations, in: Proceedings of the 2013 Conference of the North
Chapter 6. Word embeddings for applications in geosciences

American Chapter of the Association for Computational Linguistics: Human Language Technologies, 746–751.


conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL).


Chapter 7

3D lithological mapping of bore descriptions using word embeddings

Abstract

In recent years the exponential growth in digital data and the expansion of machine learning have fostered the development of new applications in geosciences. Natural Language Processing (NLP) tackles various issues that arise from human language use. In this study, NLP is applied to classify and map lithological descriptions in a three dimensional space. The data originates from the Australian Groundwater Explorer dataset of the Bureau of Meteorology, which contains the description and geolocation of bores drilled in New South Wales (NSW), Australia. A GloVe model trained with scientific journal articles and Wikipedia contents related to geosciences was used to obtain embeddings (vectors) from borehole descriptions. In parallel, and as a baseline, the descriptions were classified combining regular expressions and expert criterion. The description embeddings were subsequently classified using a multilayer perceptron neural network (MLP). The performance was evaluated using different accuracy metrics. The embeddings were triangulated and then the resulting embeddings were classified using the trained MLP and compared against a nearest neighbour (NN) interpolation of lithological classes. The mapping of the descriptions was carried out by using 3D voxels. Results show that coupling NLP with supervised classification alternatives and interpolation methods indicates a reasonable 3D representation of lithologies. This methodology is a first step in demonstrating the applicability of NLP to the geosciences, which also allows for an uncertainty quantification in the different steps of the process, such as classification and interpolation. Interpolation techniques, although acceptable, might be replaced by machine learning techniques to improve the performance of 3D models.
Statement of Contribution of Co-Authors

This chapter has been written as a journal article. The authors listed below have certified that:

6. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
7. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
8. there are no other authors of the publication according to these criteria;
9. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
10. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for this publication is:

Fuentes, I., Padarian, J., Iwanaga, T., Vervoort, W.R. 2019. 3D lithological mapping of bore descriptions using word embeddings. Geoscientific model development
7.1. Introduction

The digital era of the last decades has led to an exponential growth of information (Maarala et al., 2015). This has resulted in a change from analog to digital sources of information, which has been accompanied by increases in the storage capacity and computation power (Hilbert and López, 2011). Geosciences is one of the interdisciplinary fields of sciences that has suddenly changed since the beginning of the digital revolution. From being a poor data field it became a rich data field, integrating different sources of information such as remote sensing and geophysical surveys (Karpatne et al., 2017; Nativi et al., 2015).

Currently, sub-disciplines of geosciences, like geology, are reaching a stage of synthesis in relation to the information gathered. As pointed out by Culshaw (2005), data collection from field surveys is in decline. However, a new stage is emerging, in which new technologies allow the digitalization, storage, processing, synthesis and analysis of big (legacy) datasets.

The increasing amount of digital data has not only promoted the development of the different branches of science, but has also led to the formation of an entire new subfield of sciences, whose sole purpose is the detection of patterns in the data to solve problems (Boulton, 2018). This field, known as machine learning, has been widely applied to overcome difficulties caused by the use of big data.

While machine learning techniques have been used extensively in geosciences (Smirnoff et al., 2008; O’Brien et al., 2015; Lary et al., 2016), Natural Language Processing (NLP), which includes handling and analysing the relationships between words (Nadkarni et al., 2011; Jain et al., 2018), has seldom been applied. This is caused by a bias of scientific knowledge towards numerical data (McBratney et al., 2018).

However, large amounts of geological and pedological information have been recorded as descriptions to categorise the materials and provide qualitative information to scientists. Neglecting this information due to the above mentioned bias lacks practicality. Furthermore, the advances in NLP and machine learning mean that the subjectivity and ambiguity introduced by language might be removed by text processing and analysis (Escudero, 2006; Recasens et al., 2013).

From the most common uses of NLP, dimensionality reduction, classification, and clustering of text are the most important, which have mainly been applied to advertising and the analysis of social media (Aggarwal and Zhai, 2012). More recently, text mining tasks have expanded to other research areas, including the fields of medicine and psychiatry, and it is expected to expand to many different
science fields that deal with textual descriptions of reality (Pestian et al., 2010; Perlis et al., 2011).

Geologic datasets contain many textual descriptions. For example borehole drill descriptions which, apart from geospatial information, also contain textual information of the characteristics of the underlying materials. Due to the quantity of wells drilled, bore logs are usually one of the main data sources used for the synthesis of geological information in geologic models (Kaufmann and Martin, 2008). However, it is difficult to classify the amount of heterogeneous lithological descriptions which tend to be contained in short sentences, and are highly focused towards geologists. The semantic analysis of these descriptions is therefore constrained by the specialized geological lexicon. This makes NLP an interesting alternative to test the classification of bore log descriptions for the development of 3D geologic models.

The main difficulty of text mining is around the way in which text can be processed and analysed. It is clear that a collection of characters - referred to as strings - which make up the textual framework, cannot be used in isolation. Instead, NLP is used to pre-process and transform text into a numerical or network representation (Srivastava and Sahami, 2009). Several representation schemes have been proposed, but these have been mostly focussed on general domain text (Mikolov et al., 2013; Pennington et al., 2014; Bojanowski et al., 2017), which usually perform poorly in domain-specific tasks. Padarian and Fuentes (2019) generated word embeddings specifically trained for geosciences, which will be used in this study.

The main objective of this study is the production of three dimensional lithological maps developed by applying NLP techniques on a big dataset with borehole drilling descriptions.

7.2. Materials and methods
    7.2.1. Study area

The focus of the study is New South Wales (NSW), Australia. NSW is one of the 6 states of Australia geologically characterized by several sedimentary basins, overlying the Ordovician to Early Cretaceous basement (Figure 7.1). In the eastern part of Australia, the presence of some orogens and fold belts caused by different deformation events are the main source of sediments that fill the basins (O’Neill and Danis, 2013; Welsh et al., 2014).
Figure 7.1. Study area and lithologic framework depicting the different orogens that are the main source of sediments filling the sedimentary basins in NSW. Different colours represent the different surface lithologies.

7.2.2. Data and pre-processing steps

The main source of lithological data used in this study is the groundwater database obtained from the Australian Groundwater Explorer of the Bureau of Meteorology. It contains the geolocation and the bore logs of all the boreholes drilled in NSW. The dataset contains 100,582 boreholes (Figure 7.2) and 835,411 descriptions. Each borehole can have several descriptions associated to the different underlying lithologies. In addition, each description in the dataset is classified into a MajorLithCode, which corresponds to the lithological classification in the dataset.
There are 549 different lithologies in the dataset, several of which aggregate descriptions into ordinal categories, or aggregate descriptions based on irrelevant adjectives, such as colour or weathering state. Another dataset characteristic is the heterogeneity of descriptions and the unbalance between lithological classes. Thus, 10 classes with multiple descriptions contain around 82% of the dataset descriptions, and the distribution of the length of sentences within the descriptions is highly right skewed (Figure 7.3).
Figure 7.3. Frequency of the descriptions length in the dataset used.

The descriptions contained in the bore logs were pre-processed such that:

- The set of descriptions was tokenized (divided into words) and lemmatization was applied to all nouns. This simply involves removing inflections at the end of words in order to get the “lemma” or root of the words. In this step lists of tokens were obtained.
- All tokens with non-alphabetic characters and tokens with less than 3 characters were removed.
- The remaining tokens were converted to its lowercase form.
- Stopwords (a set of words frequently used in language which are irrelevant for text mining purposes) were removed.

In order to obtain a vectorial (numerical) representation of the words (embedding), the GeoVec model developed by Padarian and Fuentes (2019) was applied to the words from the descriptions. The model consist of an application of the GloVe model (Pennington et al., 2014) to the geosciences domain, and was trained with a corpus of over 300,000 scientific full-text articles and over 1,000 Wikipedia articles associated to geosciences.

Since the unsupervised model relies on a matrix of co-occurrence between words, the distance and angles between words in the vectorial space generated from the corpus identify different relationships. Figure 7.4 shows examples of how GeoVec depicts the semantic relationship between words, linking either specific lithologies to their corresponding rock type (Figure 7.4 left panel), or minerals and their corresponding mineral group (Figure 7.4 right panel). These groups can then be used to evaluate the numerical vectorial representation of words, which is an “embedding”.

Another property that can be obtained from this vectorial space is the interpolation between concepts. This means words between concepts can be found, which may have an empirical meaning. In Figure 7.5 (left) it can be observed how different particle sizes are found and sorted between the “clay” and “boulder” words, being the scale of increasing particle sizes: clay<silt<sand<gravel<cobble<boulder. In Figure 7.5 right a scale of metamorphic grade was found between two extreme terms “slate” and “migmatite”. In this paper we further explore this type of word embedding interpolation with an application to 3D spatial modelling.
Figure 7.5. Set of interpolation tokens obtained between two extreme terms showing scales of particle size (left) and metamorphic grade (right; Padarian and Fuentes, 2019).

As further supporting data, the Shuttle Radar Topography Mission (SRTM) digital elevation data was obtained for NSW at 30 m resolution. It was used in the mapping step to clip the interpolated lithologies to the terrain surface.

Using the borehole dataset a bore density map was developed prior to the interpolation and mapping of the lithological classes (Figure 7.6). Based on the sectors with high bore density, two areas of interest (AOI) encompassing the towns of Moree and Coleambally were selected to develop lithological 3D models because of the spatial heterogeneity of wells.
Chapter 7. Lithological 3D mapping

Figure 7.6. Bore density map and selected areas of interest (AOIs). The zoomed satellite images correspond to false colour (bands 3-4-1) Landsat 5 images.

The model, machine learning techniques and text processing tools used in this study were developed through the GloVe (Pennington et al., 2014), scikit-learn (sklearn; Pedregosa et al., 2011), and the natural language toolkit (NLTK; Bird et al., 2009) libraries implemented in the Python programming language.

7.2.3. Description embeddings and classification

To create a numerical representation for each description, an average of the constituting word embeddings was calculated, yielding a single vector of length 300. This is a simple and commonly used alternative to create representative texts (Pagliardini et al., 2018).

For the supervised classification of embeddings, a semi-manual classification of the descriptions, combining the use of regular expressions and expert criteria, was
carried out on a subset of over 700,000 points. This consisted in a two-step procedure. In the manual first step, lithologies were assigned based on the descriptions, while the second step corresponded to an aggregation of specific lithologies into a series of major lithological groups (Table 7.1) based on our interpretation of the detailed lithologies. For instance, intrusive igneous rocks are aggregated into a single class, whilst sedimentary rocks are divided in several major groups such as sandstones, shales, and limestones. The same applies for sediments, where two groups are identified based on the granulometry of sediments.

### Table 7.1. Aggregation of lithologies into Major classes.

<table>
<thead>
<tr>
<th>Major Lithology</th>
<th>Lithologies included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volcanic</td>
<td>basalt; volcanic; lava; tuff; breccia; rhyolite; agglomerate; ignimbrite; zeolite; andesite; latite; trachyte; scoria; dacite; pyroclastic</td>
</tr>
<tr>
<td>Intrusive</td>
<td>granite; diorite; porphyr; dolerite; igneous; feldspar; granodiorite; syenite; monzonite; pyroxenite; quartz</td>
</tr>
<tr>
<td>Metamorphic</td>
<td>slate; phyllite; schist; soapstone; gneiss; serpentine; mica; amphibolite; hornfels; pegmatite; metamorphic; marble; quartzite; biotite</td>
</tr>
<tr>
<td>Sandstone</td>
<td>sandstone; greywacke; arkose; wacke</td>
</tr>
<tr>
<td>Conglomerate</td>
<td>conglomerate</td>
</tr>
<tr>
<td>Shale</td>
<td>mudstone; claystone; siltstone; shale; argillite</td>
</tr>
<tr>
<td>Limestone</td>
<td>limestone; dolomite; calcite; siderite; chalk; marl; calcite</td>
</tr>
<tr>
<td>Carbonaceous</td>
<td>carbonaceous; coal; lignite; wood; charcoal; bitumen</td>
</tr>
<tr>
<td>Chemical</td>
<td>silcrete; laterite; basite; ironstone; cement; chert; jasper; gypsum; apatite; pyrite; opal</td>
</tr>
<tr>
<td>Soil</td>
<td>soil; topsoil; subsoil; earth;</td>
</tr>
<tr>
<td>Fine sediments</td>
<td>clay; mud; pug; bentonite; kaolinite; silty clay; loam; sandy loam; silty loam; clay loam; sandy clay loam; drift; stones clay; clay gravel; mud gravel; clay boulders; silty; sandy silt; silty gravel; gritty clay; sandy clay; silty sandy clay; mud sand; clay sand; silty clay sand</td>
</tr>
<tr>
<td>Coarse sediments</td>
<td>sand; silty sand; gravel; stones gravel; stones sand; sand gravel; sand boulders; clay sand gravel; pebbles, boulders, stones; blue metal</td>
</tr>
<tr>
<td>Bedrock</td>
<td>bedrock</td>
</tr>
<tr>
<td>Alluvium</td>
<td>alluvium</td>
</tr>
<tr>
<td>Sedimentary</td>
<td>sedimentary</td>
</tr>
<tr>
<td>Water</td>
<td>water</td>
</tr>
<tr>
<td>Cavity</td>
<td>cavity</td>
</tr>
<tr>
<td>Peat</td>
<td>peat</td>
</tr>
</tbody>
</table>
While the presented grouping (Table 7.1) is qualitative, it captures the major variation in the original dataset. In the future, a natural language processing algorithm could be envisioned that would do this classification.

One of the difficulties of the semi-manual classification is the high number of descriptions and the ambiguity of some. These can lead to the misclassification of some of the descriptions, which must be considered when evaluating any supervised classification. For example the description “Sand some clay trace gravel brown damp loose medium poorly sorted” was classified as “Coarse sediments”. However, the description is ambiguous, which can be further explored by using description embeddings.

For the supervised classification, a stratified random sampling was implemented such that a 10% of the dataset was included into a test subset. The remaining 90% of the dataset was subsequently divided randomly into a training and a validation subsets, which accounted for the 90% and 10% of the remaining subset, respectively.

A Multi-layer perceptron (MLP) neural network was trained to classify the resulting embeddings. It consisted of 3 fully-connected layers with 100 neurons each (Figure 7.7). MLP networks have been proven to be effective in classification tasks (Gardner and Dorling, 1998), taking advantage of all the combinations of features of the layer sequences. The network was trained using the Adam optimizer (Kingma and Ba, 2014), with 30 epochs (times that the neural network passes through the entire training data) and a batch size (training samples simultaneously propagated through the network during training) of 100. The output layer of the network yields the probability of the embeddings belonging to the different lithological classes found in the dataset, which, unlike the semi-manual classification, allows evaluation of the ambiguity of the descriptions.
Three metrics were used to evaluate the performance of the MLP classification, considering that the classes in the dataset are unbalanced. For the classification assessment, lithological classes with the higher probabilities from the classifier were used. The performance metrics included the accuracy, f1 (weighted mean of precision and recall), and balanced accuracy scores. All methods range from 0 to 1, with 1 being the exact match. In addition, a confusion matrix was constructed to evaluate the classification for the different lithological classes (Congalton, 1991).

One of the advantages of using embeddings in conjunction with the MLP classifier is that, in contrast to a manual classification, it allows quantification of the uncertainty of the classification. In this case, the normalised Shannon entropy was used as a measure of uncertainty using Eq. 7.1 (Saco et al., 2010):

$$H[P] = \left[ -\frac{\sum_{i=1}^{n} p_i \ln p_i}{S_{\text{max}}} \right]$$  \hspace{1cm} (7.1)

where $n$ corresponds to the number of classes, $p_i$ is the probability of each class, and $S_{\text{max}}$ is equal to $\ln n$. In this case, 0 represents a value of null uncertainty in the classification, and 1 a very high uncertainty.

Additionally, an assessment of the ambiguity of the classification was also carried out based on an estimated Confusion Index ($CI$; Burrough et al., 1997) by selecting the two predicted classes with highest probabilities at each point as shown in Eq. 7.2:

$$CI = \left[ 1 - (\mu_{\text{max}_i} - \mu_{(\text{max}-1)_i}) \right]$$  \hspace{1cm} (7.2)
where $\mu_{\text{max} \ i}$ corresponds to the probability of the maximum likelihood class at site $i$, while $\mu_{(\text{max} - 1) i}$ is the value of the second largest likelihood membership at site $i$. $CI$ values range from 0 to 1, being 0 a value of null ambiguity in the classification, and 1 an index that implies high confusion, and therefore, high ambiguity.

### 7.2.4. Interpolation and mapping of classes/embeddings

Instead of using a regular grid sampling design, the lithologies (classes and vectors) were extracted based on the Gallerini and Donatis (2009) methodology, which considers a maximum depth interval to extract lithological sample points, depending on the thickness of the strata or layers. This approach was used because it more accurately represents stratigraphic sequences, without missing geological information in thin strata.

Since the generation of 3D models requires continuous spatial data, the embeddings need to be interpolated. Therefore, a Delaunay triangulation followed by a barycentric linear interpolation of the embeddings (referred to as triangulation in the following sections) was carried out. The interpolated vectors were subsequently classified using the trained MLP classifier. In order to compare the interpolation of the embeddings, a nearest neighbour (NN) interpolator, which is commonly used for interpolation of categorical data (Baboo and Devi, 2010; Babak, 2013; Li and Heap, 2014), was applied to the lithological classes from the semi-manual classification.

The interpolations were performed sequentially in depth through a 2.5D approach by layering the lithologies in the dataset using a depth interval of 1 m (Falivene et al., 2007). Training and validation of the interpolations were done by splitting the bores in the dataset into test, training and validation bores in the same proportion described in the classification step. The interpolations were trained using the training dataset, and were evaluated using the f1, accuracy and balanced accuracy scores. The NN interpolation gives directly interpolated lithological classes, while the triangulation retrieves interpolated embeddings, which were passed to the MLP classifier to obtain the final lithological classes using NLP. A mapping of the interpolated lithologies was finally carried out using voxels in a 3D space.

Uncertainty maps were obtained through bootstrapping, using 100 iterations. In the NN interpolation, a lithological class probability was obtained for each class in each voxel by counting its occurrences and dividing it by the 100 iterations. In the triangulation, the probabilities obtained from running the classifier were averaged in each voxel. The uncertainties for both methodologies were then estimated through the $CI$ and the normalised Shannon entropy as described in section 7.2.3.
Chapter 7. Lithological 3D mapping

7.3. Results

7.3.1. Classification performance

The confusion matrix from the MLP classifier (Table 7.2) indicates the unbalance in the different lithological classes in the borehole descriptions, where alluvial sediments (coarse and fine grained) dominate the NSW landscape. These are followed by sedimentary lithofacies, with a predominance of sandstones, followed by shales.

Table 7.2. Confusion matrix generated comparing the semi manual and the automated classification using description embeddings.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>829</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>1</td>
<td>30</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>887</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>527</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>19</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>577</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>1</td>
<td>3,546</td>
<td>0</td>
<td>1</td>
<td>81</td>
<td>2</td>
<td>235</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>32</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3,925</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>81</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2,106</td>
<td>54</td>
<td>17</td>
<td>112</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>0</td>
<td>70</td>
<td>0</td>
<td>22</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>2,412</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>92</td>
<td>0</td>
<td>12</td>
<td>133</td>
<td>608</td>
<td>73</td>
<td>6,898</td>
<td>78</td>
<td>22</td>
<td>56</td>
<td>1</td>
<td>693</td>
<td>0</td>
<td>265</td>
<td>121</td>
<td>274</td>
<td>41</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>124</td>
<td>2,702</td>
<td>143</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>71</td>
<td>0</td>
<td>27</td>
<td>6</td>
<td>13</td>
<td>0</td>
<td>3,092</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
<td>92</td>
<td>1</td>
<td>13</td>
<td>6,396</td>
<td>76</td>
<td>304,999</td>
<td>105</td>
<td>54</td>
<td>93</td>
<td>4</td>
<td>780</td>
<td>3</td>
<td>453</td>
<td>604</td>
<td>305</td>
<td>2</td>
<td>313,976</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>16</td>
<td>49</td>
<td>3</td>
<td>179</td>
<td>18,659</td>
<td>1</td>
<td>47</td>
<td>0</td>
<td>42</td>
<td>2</td>
<td>51</td>
<td>2</td>
<td>65</td>
<td>1</td>
<td>19,119</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>16</td>
<td>1</td>
<td>150</td>
<td>3</td>
<td>2,687</td>
<td>6</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>22</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>2,912</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>83</td>
<td>20</td>
<td>0</td>
<td>7,760</td>
<td>0</td>
<td>41</td>
<td>2</td>
<td>148</td>
<td>28</td>
<td>172</td>
<td>2</td>
<td>8,277</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>6</td>
<td>47</td>
<td>2</td>
<td>14</td>
<td>146</td>
<td>15</td>
<td>417</td>
<td>73</td>
<td>4</td>
<td>20</td>
<td>0</td>
<td>71,820</td>
<td>3</td>
<td>603</td>
<td>1</td>
<td>49</td>
<td>30</td>
<td>73,250</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>123</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>3</td>
<td>105</td>
<td>0</td>
<td>1</td>
<td>96</td>
<td>4</td>
<td>1,510</td>
<td>61</td>
<td>11</td>
<td>38</td>
<td>0</td>
<td>1,122</td>
<td>1</td>
<td>79,804</td>
<td>19</td>
<td>110</td>
<td>57</td>
<td>82,942</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>63</td>
<td>1</td>
<td>255</td>
<td>13</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>12</td>
<td>43,135</td>
<td>14</td>
<td>0</td>
<td>43,510</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>10</td>
<td>145</td>
<td>15</td>
<td>424</td>
<td>33</td>
<td>1</td>
<td>20</td>
<td>0</td>
<td>52</td>
<td>4</td>
<td>116</td>
<td>9</td>
<td>29,870</td>
<td>71</td>
<td>30,776</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>393</td>
<td>447</td>
</tr>
</tbody>
</table>

Total 834 549 3,893 88 2,176 142,822 9,912 315,460 19,065 2,795 8,068 31 74,767 138 81,562 43,932 30,895 600 310,587


Lithological classes that had lower accuracies were those with a fewer number of descriptions. The class water had the worst classification, and it was mostly misclassified as volcanic, shale and sandstone. The misclassification indicates the main water bearing lithologies, or in some other cases indicates overlap between classes. For instance, bedrock is a broad concept that can refer to different lithologies, but had no further lithological detail in the descriptions. Another example is the soil and fine sediments classes. Most soils are composed of a mix of different grain sizes mineral particles and organic matter, but can also be described as the textural class of those materials, which are classified as fine grained sediments for
the classifier instead of soils. This leads to a confusion that if not reflected in the classification step, will affect the interpolation results.

In addition, as mentioned in the methodology, the semi-manual classification also created misclassification due to the ambiguity of descriptions. In some cases, semi-manually misclassified descriptions were correctly classified by the MLP classifier, but not reflected in the metrics used, since the semi-manual classification is used as reference. For instance, in the aforementioned example the MLP classified the description “Sand some clay trace gravel brown damp loose medium poorly sorted” as “Fine sediments”, which includes the “sand clay” lithology.

Overall, the MLP classification with the averaged word embeddings achieved good results based on the given inputs, with 0.958 accuracy and a balanced accuracy index of 0.864. Based on this overall performance (shown in Table 7.3) the MLP classifier was used for subsequent statistical and spatial analysis.

### Table 7.3. Classification performance of MLP neural network for the sentence embeddings obtained by averaging the word embeddings of the descriptions.

<table>
<thead>
<tr>
<th>Embeddings</th>
<th>Accuracy</th>
<th>F1</th>
<th>Balanced accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>0.965</td>
<td>0.937</td>
<td>0.939</td>
</tr>
<tr>
<td>Validation</td>
<td>0.958</td>
<td>0.868</td>
<td>0.864</td>
</tr>
</tbody>
</table>

Since the semi-manual classification is a time-consuming task, the training performance as a function of the size of the training dataset was further explored (Figure 7.8). As expected, more training samples lead to higher accuracy of the validation and a slight decrease in the training accuracy, which results in minimum error between the curves at around 500,000 samples. Regardless of the increase in accuracy, even with a small fraction of the dataset used for training, the classification accuracy is greater than 0.9. This means that over 50,000 training samples, the classifier leads to good results, independent of any further increase in training sample size. This means a threshold in the resources used for the semi-manual classification can be based on this training size.
Figure 7. 8. Training size and its effect on the classification accuracies for the training and validation datasets and on the classifier training time.

7.3.2. Classification uncertainty

Analysing the mean normalised entropy per lithological class and the lithological class percentages in the different AOIs, it is clear that the highest entropies are associated with lithologies with a small occurrence, and except for the water and sedimentary classes, the entropies are mainly low (0-0.1; Figure 7.9). Again, sediments dominate because NSW is mainly composed of sedimentary basins and the boreholes are mainly drilled in alluvial deposits (for water supply). Obviously the distribution of lithologies has a higher spread for the entire NSW dataset, compared with the two AOIs.
Chapter 7. Lithological 3D mapping

Figure 7.9. Mean normalised Shannon entropy per lithological class (left panel) and the percentage of each lithological class (right panel) in the different study areas.

7.3.3. Interpolation and 3D mapping

An overall assessment of the two interpolation methods used in both AOIs is in Table 7.4. While both interpolation methods have acceptable results, the Coleambally AOI, on average, has better performance than Moree.
Table 7.4. Mean performance metrics at Moree up to 60 m depth, and in Coleambally up to 80 m depth.

<table>
<thead>
<tr>
<th>AOI</th>
<th>Interpolation</th>
<th>Accuracy</th>
<th>F1</th>
<th>Balanced accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NN training</td>
<td>0.984</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>Moree</td>
<td>NN validation</td>
<td>0.600</td>
<td>0.415</td>
<td>0.487</td>
</tr>
<tr>
<td></td>
<td>Triangulation training</td>
<td>0.964</td>
<td>0.922</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>Triangulation validation</td>
<td>0.600</td>
<td>0.446</td>
<td>0.494</td>
</tr>
<tr>
<td>Coleambally</td>
<td>NN training</td>
<td>0.988</td>
<td>0.984</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td>NN validation</td>
<td>0.720</td>
<td>0.569</td>
<td>0.579</td>
</tr>
<tr>
<td></td>
<td>Triangulation training</td>
<td>0.961</td>
<td>0.659</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>Triangulation validation</td>
<td>0.705</td>
<td>0.564</td>
<td>0.571</td>
</tr>
</tbody>
</table>

Figure 7.10 shows the performance of the interpolation in depth (left) in relation to the number of samples (right). In both areas of study the number of training and validation points increases in the first meters depth, but diminishes rapidly at deeper depths. In Coleambally, the maximum number of sampled points reaches around 10,000, whilst in Moree there are always less than 1,000. The decrease of samples with depth is more gradual in Moree. However, after 55 m depth a sudden reduction in samples is observed, most probably due to reaching the bottom of a productive aquifer, which leads to a sudden drop in the interpolation performance.
Figure 7.10. Interpolation accuracies in depth (left) and number of points used in the interpolation (right).

Accuracies fluctuate with depth due to the 2.5D interpolation scheme used on a 1 m interval. Even though a very general trend may be detected suggesting that the interpolation of shallow lithologies leads to a better performance, this does not relate clearly with the number of training samples. Thus, while at shallow depths there is a slight performance increase, this is followed by a steady decrease with depth in Coleambally, which relates to the peak in the number of training samples (Figure 7.10). A completely different picture emerges in Moree, where surface lithologies have high accuracies, which diminish rapidly with depth. A Pearson correlation test was used to assess the relationship, indicating that both AOIs had a positive correlation between accuracies and the number of training and validation samples ($0.38 < r < 0.51$, $p$-values $< 0.05$). Therefore, the number of samples affects the performance of the interpolator, which must be considered prior to building of geologic models. However, as the relationships were not strong, other factors affecting the performance must be taken into account, such as the complexity of the lithology.

Overall there was little difference between the two interpolation methods for each AOI (Figure 7.10).

The voxel maps of the lithologies in the Moree study area based on the two different interpolation methods are in Figure 7.11. While the flexibility of the nearest
neighbour interpolation means the interpolation can go beyond the training points in the map, the results in Figure 7.11 were masked based on a convex hull border, to make comparison of the results of both interpolation methods easier. This might reduce the uncertainty of the results outside the training samples in the interpolation area.

![Figure 7.11. Moree 3D lithological maps obtained using a NN interpolation (left) and a triangulation of embeddings (right).](image)

The lithologies at Moree are mostly sequences of sediments that alternate between fine and coarse grain sizes, except for the northeast of the AOI. Here sedimentary sequences can be observed in depth that include conglomerates and sandstones. In the area, the Mehi and the Gwydir Rivers dominate the developed alluvial landscape and lead to cenozoic alluvial deposits such as the Narrabri formation (shallower and composed of fine-medium grained sediments) and Gunnedah formation (deeper and composed by coarse grained sediments), which contain the two main alluvial aquifers (Welsh et al., 2014). These are overlain by soils and fine grained sediments, yet coarse sediments are found at shallow depths in the middle of the AOI, where the rivers are located (Figure 7.6). On the north eastern and southern extremes of the AOI, surface colluvial sediments are found (Geoscience Australia, 2012). The depth of the alluvium tends to increase to the west.

In the Coleambally AOI sediments are deeper with a predominance of fine-medium grain sizes in the lithological setting, with some patches of coarse grained sediments at depth (Figure 7.12). According to Prathapar et al. (1997), the alluvial deposits in the area have a thickness of between 100 m and 200 m, increasing westward. In
the northern area a strip of surface coarse sediments can be found, which follows the main course of the Murrumbidgee River (Figure 7.6). With depth, layers and lenses of coarse sediments indicate where palaeochannels occur (Page et al., 1996). These paleochannels act as the shallow aquifers in the region (Prathapar et al., 1997). The dominant surface lithologies are part of the Shepparton Formation which forms extensive alluvial floodplains. This Formation is described as unconsolidated to poorly consolidated fine sediments, including clay or silty clay grain sizes, having lenses of polymictic sand and gravel (Geoscience Australia, 2012). The deepest extensive coarse sediment layers that can be observed in Figure 7.12 represent the Calivil Formation, which is the most transmissive strata in the region, and therefore it has been highly exploited for irrigation purposes (Page, 1994).

Figure 7.12. Coleambally 3D lithological maps obtained using a NN interpolation (left) and a triangulation of embeddings (right).

7.3.4. Uncertainty of 3D models

Clearly, the uncertainty for the Moree AOI (Figure 7.13) increases for small lithological patches surrounded by dominant lithological classes, indicating that the interpolation alternatives do not perfectly match the actual terrain distribution of rocks and sediments. In the case of uncertainties obtained through triangulation, the limit between lithologies tends to indicate high uncertainty values. The overall uncertainties using both interpolation alternatives are relatively low. The mean CI value for the NN interpolation was 0.095 while it was 0.275 using the triangulation. Mean entropies were 0.052 and 0.093 using the NN interpolation and the
triangulation, respectively. Additionally, a two-sample Kolmogorov-Smirnov test indicated that the NN interpolation (Figure 7.13 above) has significantly lower uncertainties (for both CI and entropies) than the triangulation ($p$-value < 0.05; Figure 7.13 below).

![Figure 7.13. Uncertainty mapping for the Moree AOI using the NN interpolation (above) and the triangulation of embeddings (below). Entropy scales were normalised to the range of values obtained.](image)

The uncertainty of the Coleambally lithological maps is in Figure 7.14. Again, uncertainty values are mostly low. However, some small patches of higher uncertainty can be observed. Again the NN interpolation (Figure 7.14 above) has significantly lower uncertainties ($p$-value < 0.05; mean CI of 0.069 and mean entropy of 0.037) than the triangulation (mean CI of 0.248 and mean entropy of 0.093; Figure 7.14 below), yet in both cases the uncertainties are lower than in the Moree AOI.
Figure 7.14. Uncertainty mapping for the Coleambally AOI using a NN interpolation (above) and a triangulation of embeddings (below).

The distribution of the mean confusion index ($CI$) also fluctuates with depth (Appendix). A moderate negative correlation ($-0.44 \leq r \leq -0.55; p$-values < 0.05) was found between CIs and triangulation accuracies in depth for both AOIs, which means that the ambiguity of embeddings also affects the interpolation performance.

7.4. Discussion

There are limited studies applying text mining techniques to geosciences (Padarian and Fuentes, 2019). For example, Pollock et al (2012) used regular expressions and geospatial data to map lithological features by using the interpolation of match scores (from the regular expressions). However, this had several limitations, such
as a small area of application, the use of only three different lithologies, and the lack of automation in the methodology.

In this study, we move a step further in the use of text mining to address geosciences problems. By combining NLP (with the simplest conversion from word embeddings to sentence embeddings), a supervised classification method (MLP neural network) and the simplest of interpolation techniques (a triangulation of embeddings) we were able to build 3D lithological maps from borehole descriptions, and assess prediction uncertainties.

The use of word embeddings led to very accurate results in the classification step compared with a semi-manual classification. However, since the description embeddings are built up based on a specialised lexicon, other alternatives rather than a simple averaging of words embeddings, such as sequential denoising autoencoders or neural networks with complex architectures (LSTM) might lead to an improvement of the classification (Wieting et al., 2015).

Overall, the results obtained were comparable with those derived using a semi-manual expert criteria classification, and even slightly better for the interpolation results in Moree, and is therefore suggested to be useful for such applications. Another advantage of the automated method is in the performance of the classification step. Even a small dataset (≈50,000 samples; Figure 7.8) leads to relatively good performance.

Even though the current implementation is not fully automated, it can be considered as a first step in the automation of these tasks, and a first step demonstrating the applicability of NLP to the geosciences. For a full automation, unsupervised classification alternatives will need to be coupled to NLP. This is currently part of our future research.

Most geological studies that address three dimensional modelling using borehole and other data sources do not provide an evaluation of accuracies, but generally assume a good performance (Kaufmann and Martin, 2008; Kessler et al., 2009; Høyer et al., 2015). For instance, Gallerini and Donatis (2009) did not carry out an evaluation of their model performance arguing that fluvial environments present heavy facies variations, and therefore, used the entire dataset in the model. Even though this argument seems reasonable, this precludes evaluation of the performance of the final model. However, in some studies that do present performance metrics (Falivene et al., 2007; He et al., 2010; Hengl et al., 2014), 3D models usually have limitations due to the moderate performance of interpolation techniques in the 3D space. As pointed out by Kumar et al. (2000), the interpolation between boreholes is a hard task for humans and leads to only moderate reliability.

This study suggests acceptable results for the presented 3D lithological models. However, the 2.5D modelling scheme precludes inclusion of vertical information in
Chapter 7. Lithological 3D mapping

the interpolation. An interpolation in a 3D space does not necessarily lead to better results (Wu et al., 2005), which may be related to the variance of the data (Sahlin et al., 2014), and the type of lithological classes. Abrupt transitions between different sediments/facies can be accurately captured by the use of 2.5D schemes, while smooth transitions might be better represented using 3D schemes. In this case, the 2.5D scheme led to slight fluctuations in the interpolation performance in depth, which depended moderately on the number of samples and the ambiguity of embeddings, and possibly on the complexity of the lithological setting. Since traditional interpolation methods show a moderate performance when working on 3D arrays (Zhou et al., 2005; Sahlin et al., 2014; Hengl et al., 2014), this opens the door for the use of machine learning techniques to address these tasks.

Uncertainty is usually discussed in different 3D geologic modelling studies; however, most of them avoid its quantification (Wu et al., 2005; Gallerini and Donatis, 2009; Zhu et al., 2012). Even though both methodologies produce 3D models that are not a perfect representation of reality (Table 7.4), only a proper uncertainty analysis is possible to identify where the performance of models is unsatisfactory (Lindsay et al., 2012). Compared to the semi-manual classification and the NN interpolation, the 3D lithological models derived from word embeddings show a higher uncertainty, mostly due to the cumulative effect of uncertainties obtained through the classification and interpolation stages, which yields a more realistic representation (Jones et al., 2004). Thus, the proposed thorough evaluation and mapping of uncertainty can be used to guide future explorations for more accurate results, which is a clear advantage of using word embeddings.

While several geologic software packages have been developed to build 3D models (RockWare®, Leapfrog®, Georeka, GSI3D), most of them have a high cost. In this case, using open source modules implemented in Python allowed the development of 3D lithological maps using voxels, capturing the lithological setting of large areas.

Different applications can be found for the resulting 3D lithological models. They can be included as input in more complex geological models, and these can also be used in hydrogeological modelling, ore prospecting, territorial ordering and environmental studies, indicating valuable data synthesis for the geosciences field.

7.5. Conclusions

Applying NLP is useful for geoscience applications. By integrating NLP, machine learning algorithms and spatial interpolation techniques, lithological 3D maps could be obtained. NLP and machine learning can automate 3D geological mapping from text input bore descriptions, which might be further explored using unsupervised classification algorithms. This allows the otherwise qualitative and manually
interpreted data to be applied quantitatively, opening a new information source for geoscience applications.

Firstly the classification of lithological description embeddings through MLP neural networks is very accurate and just a small fraction of samples used for training the classifier results in high accuracy. Secondly, NLP allows for an uncertainty quantification in the different stages of the 3D model generation. Simple interpolation techniques using a 2.5D approach give an acceptable performance, demonstrating that the triangulation of embeddings and their subsequent classification gives equivalent results, and in some cases even slightly better, than by using a NN interpolation of lithological classes obtained manually.

The 3D lithological models generated using voxels through Python public libraries correspond to reasonable representations of complex lithological settings in relatively large areas, which can be further improved using uncertainty maps.

### 7.6. Appendix

![Confusion indices](image)

**Figure 7.A 1.** Mean and standard deviation of confusion indices with depth for Moree (left) and Coleambally (right).
Chapter 7. Lithological 3D mapping

Figure 7.A 2. Confusion index histograms for the Moree (left) and Coleambally (right) interest areas until 100 m depth.

7.7. References


Chapter 7. Lithological 3D mapping


O’Neill, C., Danis, C., 2013. The geology of NSW. The geological characteristics and history of NSW with a focus on coal seam gas (CSG) resources. A report commissioned for the NSW Chief Scientist’s Office, Macquarie University, Sydney, Australia.


Chapter 7. Lithological 3D mapping


the Environment, Bureau of Meteorology, CSIRO and Geoscience Australia, Australia.


Chapter 8

Site suitability and water availability for a managed aquifer recharge project in the Namoi basin, Australia

Abstract

This study was conducted in the Namoi catchment, Australia. Managed aquifer recharge (MAR) site suitability has been widely studied through multicriteria decision analysis (MCDA). However, the selection of areas for MAR project implementation can be a vague process with different validation approaches. The aim of this study was to create a site suitability map for MAR projects and conduct a sensitivity analysis for the selection of an area of interest (AOI) by combining highly suitable and low sensitive areas. Ten hydrologic and hydrogeologic criteria were chosen for the selection of sites for the MAR project. All criteria were reclassified and used in a MCDA that combined Analytic Hierarchy Processes (AHP) and pairwise comparisons to construct a site suitability map. Validation of the map and the AOI selected was performed using hydrograph data. Based on the AOI map, water availability and frequency were analysed using gauging station data. The selected AOI represents high spatio-temporal variability in natural recharge rates, highly dependent on the Namoi River streamflow. Moreover, all recharge rates in the AOI are high, particularly in the paleochannel surrounding the current Namoi River. The selected AOI coincides with an area of thick coarse sediments underlying the riverbed and demonstrates the usefulness of the proposed methodology.
Statement of Contribution of Co-Authors

This chapter has been written as a journal article. The authors listed below have certified that:

11. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
12. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
13. there are no other authors of the publication according to these criteria;
14. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
15. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for this publication is:


<table>
<thead>
<tr>
<th>Contributors</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ignacio Fuentes</strong></td>
<td>Original idea; Coding; Data analysis; Writing</td>
</tr>
<tr>
<td><strong>R. Willem Vervoort</strong></td>
<td>Supervision; Edition; Suggestions</td>
</tr>
</tbody>
</table>
8.1. Introduction

Several studies have highlighted the depletion of global groundwater resources due to unsustainable exploitation rates (Gleeson et al., 2012; Rodell et al., 2009; Döll et al., 2014). Moreover, many countries are now clearly affected by climate change, with increasing temperatures and a reduction in water storage opportunities due to either an increase in rainfall intensity or a decrease in rainfall frequency (Collins et al., 2013). The increase in temperature is also likely to increase potential evaporation in some areas (Moratiel et al., 2010), which implies a higher pressure on water resources.

Therefore, several management practices capable of supplying water demand without negatively impacting groundwater have been proposed and implemented (Rockström and Falkenmark, 2015; Harb et al., 2016; Kim et al., 2017). Most of these, carried out at regional scales, involve the building of reservoirs and the development of farm dams and ring tanks at smaller scales (Poff and Olden, 2017). However, these surface works also lead to significant water losses. For instance, in Australia, lost water in these structures can reach up to 40% of their storage capacity every year, exceeding several thousand GL yr$^{-1}$, which diminishes the water use efficiency and represents considerable economic losses (Craig et al., 2005). Among the alternative water harvesting methods, managed aquifer recharge (MAR) has gained importance. It allows replenishing or maintaining groundwater levels in times of surface water surplus, while providing a storage of water that can be used to sustain ecological functioning and irrigation needs during droughts (Rawluk et al., 2013), without leading to higher water losses by evapotranspiration (Dillon and Arshad, 2016). This method, balances and recovers groundwater resources by enhanced infiltration through in-channel modifications, floodwater spreading systems or injection wells and enables cyclical water management by storing surplus surface water in aquifers during wet years which can be pumped for use in dry years (Rawluk et al., 2013; Bouwer, 2002). However, some disadvantages of MAR have also been reported. This includes a limit on the recharge capacity depending on aquifer characteristics such as storage and transmission, the high costs of implementation and operation, especially using injection systems, requiring high energy consumption, and the potential of losses due to mixing when brackish aquifers are used (Dillon et al., 2009).

Additionally, before a MAR project can be implemented, studies are needed to assess its feasibility (Ringleb et al., 2016). In particular, the selection of suitable sites for the location of a project is not a trivial task, because it requires the combination and prioritisation of different criteria. The most common technique for the site suitability
assessments is a multicriteria decision analysis (MCDA) which can be easily combined with the use of geographical information systems (GIS; Malczewski and Riner, 2016). The different techniques applied in the MCDA seek to assign weights and combine different GIS criteria based on the study goal (Sallwey et al., 2018).

Several studies highlight the usefulness of MCDA in context of data scarcity (Giove et al., 2009; Paquette and Lowry, 2012; Paul et al., 2016; Rousseau et al., 2017). While modeling algorithms and its performance depend on the quantity and quality of the datasets used, MCDA techniques integrate different data sources with the researchers’ judgement in the decision making process (Giove et al., 2009), which has been proven effective in different fields (Gamper et al., 2006; Russo et al., 2014; Paul et al., 2016).

Since the MCDA requires the selection and assignment of weights of defined criteria, some subjectivity is introduced in the decision-making process (Buchanan et al., 1998). Among the different MCDA alternatives, the Analytical Hierarchy Process (AHP) has been widely applied since it allows for a consistency check in the MCDA, which leads to a bias reduction in the analysis (Mu and Pereyra-Rojas, 2017). For that reason, the use of AHP has increasingly been used in MAR site suitability studies (Sallwey et al., 2018). Different new MCDA techniques have been developed in the last decade (Rezaei, 2015; Wang et al., 2016; Zhang et al., 2017), some of which have been proven better than AHP for particular cases (Kolios et al., 2016; Jozaghi et al., 2018), while other studies demonstrate opposite results (Feizizadeh et al., 2014; Dožić and Kalić, 2015). Thus, in the end, the goals of the different MCDA projects determine the usefulness or performance of the methods used (Guarini et al., 2018). In this regard, AHP seems to perform well when dealing with simple problems (Asadabadi et al., 2019), and much of the criticism concerning AHP has been caused by a misunderstanding of its functioning and applicability (Whitaker, 2007; Wang et al., 2009). The main advantages of AHP over other techniques are its simplicity and its flexibility and intuitive appeal to users, which makes it one of the most popular MCDA techniques (Giove et al., 2009). Additionally, a set of tools to perform AHP have been developed, including online applications, allowing work in multi-user projects to obtain a consensus between participants (Goepel, 2018). AHP has also proven useful compared to other approaches based on the criteria weighting process and for that reason has been coupled to other MCDA techniques, which have shown better results in the rating of alternatives (Mokhtar et al., 2015; Berdie et al., 2017). Given these reasons, the simple criteria structure of site suitability for MAR projects, and the focus on MAR site suitability, which depends on the criteria weighting rather than on a selection of alternatives, AHP was used in this particular study.

Even though the magnitude of available water for a MAR project can be integrated into the MCDA, this would generally lead to a loss of temporal information, and
therefore to a loss of the potential operation frequency. Accordingly, we propose first to perform the site selection and later, by refining the MAR suitable areas, an assessment of water availability. The project was therefore split into two complementary steps based on the information sources used (vectors/rasters and time series).

In addition, the location of areas for MAR projects from site suitability maps have been vaguely defined, where most studies finish with the creation of suitability maps (Ghayoumian et al., 2007; Chowdhury et al., 2009; Mahmoud, 2014). Also, different alternatives for validation have been proposed, some of them limited to the consistency of the MCDA technique applied (Kazakis, 2018). In this study, we propose to refine the area of interest (AOI) for MAR projects by coupling the site suitability and sensitivity analysis. Thus, areas with high suitability and low sensitivity in relation to the different scenarios may be selected. For validation, the evaluation of local natural recharge is proposed as an additional technique to evaluate the suitability results.

Therefore, the main objectives of this study are to develop a site suitability map and to define a potential AOI for MAR projects in an example catchment in Australia, using a MCDA, based on an AHP technique, coupling the results of the suitability and the sensitivity analysis, and validating the defined AOI through hydrograph analysis. Additionally, estimates of water volumes available for recharge and their corresponding frequencies were also assessed in a subsequent complementary step.

8.2. Materials and methods

8.2.1. Study area

The study was carried out in the Namoi catchment, located in New South Wales (NSW), Australia, which covers an area of approximately 42,000 km² (Figure 8.1). This catchment was selected due to the high level of agricultural and irrigation development. A significant area in the catchment is used for agriculture, with cotton being the most important irrigated crop. It is suggested that water withdrawals for irrigation have changed gaining sections of the Namoi River into losing sections (Kelly et al., 2013). Groundwater abstractions have also led to a decrease in groundwater levels in areas away from the river, which have dropped up to 30 m between 1978 and 2008 (Kelly et al., 2014). This offers the opportunity to study the feasibility of MAR projects to reduce groundwater decline and potentially increase water resources for irrigation as well as providing environmental services.
The catchment is characterised by a wide floodplain in the west, and highland areas in the east and south. The mean annual rainfall in the catchment is 800 mm, ranging from less than 500 mm to the west of the catchment, to over 1000 mm eastward, while potential evapotranspiration averages 1300 mm y\(^{-1}\) (McCallum et al., 2010). The surface has been carved out by the Namoi River and its tributaries, which have developed a continuous setting of Cainozoic fluvial and alluvial sediments. To the east of the catchment hydrogeologic units develop in Paleozoic fractured rocks (Figure 8.2). The fluvial filling sediments have been traditionally divided into three main formations that have been characterised as different aquifer units with spatially variable interconnectivity (Williamn et al., 1969; Kelly et al., 2013), of which the uppermost formation is referred to as the Narrabri Formation. This formation dates from the Pleistocene, has a general thickness of 30 m - 40 m (ranging from 10 m to 70 m), and consists mainly of clay-dominated sediments with minor channels of sand and gravel. Within it, an unconfined aquifer can be found, which is connected with the Namoi River in some reaches (Kelly et al., 2014). The shallow aquifer in the region according to this conceptual model is one of the main recharge sources to the underlying highly productive aquifers corresponding to the Gunnedah and the
Cubbaroo Formations, mainly conformed by sand and gravel sediments. However, this traditional conceptualization of the alluvial aquifer has been contested in different studies (Kelly et al., 2014; Acworth et al., 2015; Kelly et al., 2017) because it fails to explain the hydrogeochemical results obtained in the catchment. These studies have found the alluvial aquifer being a distributive fluvial system with different degrees of interconnectivity based on the sediment composition, which is recharged by floodwaters and river leakage, but also by the Great Artesian Basin (GAB) in different areas (Iverach et al., 2017).

![Figure 8.2](image.png)

**Figure 8.2.** Surface hydrogeologic units in the Namoi catchment.

Another dominant feature in the catchment is the presence of an upper clay layer of approximately 10 m thickness, characterised by low hydraulic connectivity, which can be found across almost all the alluvial deposits (Timms et al., 2011). This must be considered when studying the surface-groundwater interaction in the catchment and the potential for the development of a MAR project using surface spreading methods, which requires a high infiltration capacity of the sediments.
8.2.2. Data and preprocessing steps

Data sets used in this study are the Australian Smoothed Digital Elevation Model (DEM-S) (Gallant et al., 2011); the Australian Groundwater Explorer; the Australian Hydrological Geospatial Fabric v 2.1 (Geofabric); the Australian Land Use and Management (ALUM) for NSW; the Atlas of Australian Soils; and the Water NSW gauging stations.

A slope raster was obtained at a 30 m resolution from the DEM-S dataset. Drainage paths and major river channels were selected from the Geofabric product to generate drainage density and distance to river maps for all grid points in the catchment.

Surface hydrogeological units and aquifer hydraulic conductivity, yield and salinity maps were directly obtained from the Geofabric groundwater dataset (Bureau of Meteorology, 2012). A raster of groundwater levels was obtained from the monitoring bores included in the Groundwater Explorer (http://www.bom.gov.au/water/groundwater/) by averaging the groundwater levels, filtering out bores with mean groundwater levels greater than 35 m depth and with less than 10 observations in time, and interpolating (kriging) the level of the remaining bores.

A map of soil hydraulic conductivity classes was obtained from the Australian Soil Atlas (http://www.asris.csiro.au/themes/Atlas.html). A map of distances to agricultural areas (distance to users) was generated from the ALUM land use maps (ABARES, 2016).

Additionally, daily gauging station data from the Water NSW website (https://realtimedata.waternsw.com.au/) and rainfall time series data from the SILO website (https://legacy.longpaddock.qld.gov.au/silo/) were also used.

All criteria maps were rasterized, converted to the same coordinate system (GCS WGS84), and clipped to the extension of the Namoi basin.

8.2.3. Map reclassification and rating

Since MCDA requires all data to be standardised and converted to the same schema (Sallwey et al., 2018), all maps were reclassified and ranked as shown in Table 8.1 based on a review of commonly used classes, the data distribution, and parameter ranges. In this study, in-channel modifications and spreading methods were originally considered as potential techniques for the MAR project. Therefore, the criteria were selected and treated accordingly.
Table 8.1. Criteria and rating used in the site suitability study.

<table>
<thead>
<tr>
<th>Criteria groups</th>
<th>Criteria</th>
<th>Class</th>
<th>Range</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface</td>
<td>Slope (degrees)</td>
<td>very high</td>
<td>0-2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high</td>
<td>2-5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>moderate</td>
<td>5-10</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>low</td>
<td>10-30</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very low</td>
<td>&gt;30</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Distance to rivers</td>
<td>very high</td>
<td>50-300</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high</td>
<td>300-1000</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>moderate</td>
<td>0-50</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>low</td>
<td>1000-5000</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very low</td>
<td>&gt;5000</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Distance to users</td>
<td>very high</td>
<td>&lt;100</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high</td>
<td>100-300</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>moderate</td>
<td>300-500</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>low</td>
<td>500-1500</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very low</td>
<td>&gt;1500</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Drainage density (km km(^{-2}))</td>
<td>very high</td>
<td>&lt;1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high</td>
<td>1-2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>moderate</td>
<td>2-3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>low</td>
<td>3-4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very low</td>
<td>&gt;4</td>
<td>1</td>
</tr>
<tr>
<td>Aquifer</td>
<td>Hydrogeological unit</td>
<td>very high</td>
<td>Alluvial sediments</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high</td>
<td>Colluvial sediments</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>moderate</td>
<td>Weak sedimentary</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>low</td>
<td>Volcanic and competent sedimentary</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very low</td>
<td>Intrusive, metamorphic and very strong sedimentary</td>
<td>1</td>
</tr>
<tr>
<td>Aquifer Ks (m day(^{-1}))</td>
<td>very high</td>
<td>&gt;50</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high</td>
<td>20-50</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>moderate</td>
<td>10-20</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>low</td>
<td>1-10</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very low</td>
<td>&lt;1</td>
<td>1</td>
</tr>
<tr>
<td>Aquifer yield (L s(^{-1}))</td>
<td>very high</td>
<td>&gt;25</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high</td>
<td>15-25</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>moderate</td>
<td>9-15</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>low</td>
<td>5-9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very low</td>
<td>&lt;5</td>
<td>1</td>
</tr>
<tr>
<td>Aquifer salinity (mg L(^{-1}))</td>
<td>very high</td>
<td>0-250</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high</td>
<td>250-500</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>moderate</td>
<td>500-1500</td>
<td>3</td>
</tr>
</tbody>
</table>
Sites with low slopes have less runoff and are more suitable to promote infiltration. Therefore, lower slopes were ranked higher. Since floods are considered as the source of water for the project, and their occurrence takes place in the neighborhood of streams, a closer distance to rivers is indicative of higher suitability. However, very small distances were also penalized (ranked third) in order to avoid the flow from the recharge project back to the stream from a water table mound caused by the project (Sophocleous, 2002). Long distances from the river were ranked lower because they imply higher costs for the project, due to the need for hydraulic infrastructure to transport the water.

Drainage density, evaluated as the ratio between the length of the drainage path and the area that it covers (km km\(^{-2}\)), has been inversely related to the permeability because a denser underlying lithology, causes higher runoff, which leads to the development of a well-defined drainage network (Ghayoumian et al., 2007, Raviraj et al., 2017). Thus, drainage density was inversely ranked in equally separated classes.

Hydrogeologic units were reclassified and ranked, such that a range of different lithologies, from the alluvial channel and floodplain sediments to intrusive igneous, very strong sedimentary and metamorphic lithologies, could be observed. These were classified as having very high (alluvial sediments) to very low suitability (hard rocks) for the MAR location.

Aquifer hydraulic conductivity (Ks) and yield were reclassified and ranked favouring higher conductivity and recovery (yield) rates. Aquifer salinities were classified and ranked based on a salinity classification for irrigation water developed by the Victorian government (http://agriculture.vic.gov.au/).

Groundwater depths were ranked using two criteria. On the one hand, very shallow depths constrain the potential for the project because the recharge in those areas

<table>
<thead>
<tr>
<th>Underground Groundwater depth (m)</th>
<th>very high</th>
<th>high</th>
<th>moderate</th>
<th>low</th>
<th>very low</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>8-15</td>
<td>5-8 or 15-20</td>
<td>3-5 or 20-25</td>
<td>1-3 or 25-30</td>
<td>&lt;1 or &gt;30</td>
</tr>
<tr>
<td>very low</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Soil Ks (mm day(^{-1}))</th>
<th>fast</th>
<th>moderate</th>
<th>slow</th>
<th>very slow</th>
<th>extremely slow</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>&gt;500</td>
<td>50-500</td>
<td>5-50</td>
<td>1-5</td>
<td>Reservoirs (&lt; 1)</td>
</tr>
<tr>
<td>very low</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
can lead to groundwater discharge and water logging (Kazakis, 2018). In addition, some interaction between soil and sediment particles with the infiltrating solution, which partly depends on the vadose zone depth, is required, to reduce the potential harm of groundwater pollution (Rahman et al., 2012). On the other hand, elevated groundwater depths, based on Darcy’s law, imply longer times for water transport and a higher residence time of ponded water, which increases evaporation losses. Therefore, intermediate depths were ranked higher in this study.

Higher soil Ks was ranked higher because as it was pointed out by Arshad et al. (2014), infiltration rates are one of the key properties of the physical medium that influences the economic feasibility of a MAR project in a catchment. Since the purpose of the project is to increase the available groundwater for irrigation purposes, smaller distances to irrigation fields (distance to users) were ranked higher.

### 8.2.4. Decision criteria and site suitability map

The site suitability map to locate a project was based on a multi criteria decision analysis (MCDA) using Analytic Hierarchy Process (AHP) and pairwise comparisons. The AHP technique is one of the most commonly used MCDA tools, which applies an eigenvalue approach to the pairwise comparisons (Vaidya and Kumar, 2006), reducing the subjectivity associated with the definition of weights (Kazakis, 2018). The pairwise comparisons were carried out using the Saaty’s pairwise comparison scale between criteria pairs (Saaty, 2012), which were treated independently. The scale ranges between 1, where both criteria are equally important, and 9, where one criterion is far more important than the other. Different pairwise comparisons were carried out to obtain a consolidated decision matrix through the geometric mean of individual matrices according to eq. 8.1:

$$c_{ij} = \exp\left[\frac{1}{N}\sum_{k=1}^{N}\ln a_{ij(k)}\right]$$  \hspace{1cm} (8.1)

where $c_{ij}$ is the consolidated decision matrix, being $i$ and $j$ the rows and columns of the matrices, $k$ is each participant, $N$ the total number of participants, and $a_{ij(k)}$ corresponds to the decision matrix of each participant.

The AHP technique was evaluated through a Consistency Ratio (CR), which corresponds to the ratio between the consistency index of the pairwise comparison matrix (CI) and the consistency index of a random-like matrix, known as random index (RI), where values smaller than 0.1 are considered acceptable to continue the decision making analysis (Mu and Pereyra-Rojas, 2017).

The consistency index is estimated by eq. 8.2:
\[ CI = \frac{\lambda_{\text{max}} - n}{n - 1} \]  \hspace{1cm} (8.2)

where \( n \) is the number of compared elements (\( n \) = number of criteria) and \( \lambda_{\text{max}} \) is the largest eigenvalue of the matrix. All AHP analysis was carried out using the online AHP-OS software (https://bpmsg.com/ahp-online-system/) (Goepel, 2018).

A sensitivity analysis was also carried out, using 7 alternative weighting schemes (Table 8.2) for the criteria groups defined in Table 8.1, mainly to find areas relatively insensitive to the weighting schemas. In the first six scenarios, the weights were alternately applied to the weightings obtained from the AHP technique such that one group was excluded from the analysis, and the others split the weights into one and two thirds. In the last scheme, all criteria were equally weighted.

**Table 8.2.** Different weighting schemas for the groups of variables (Table 8.1) used in the sensitivity analysis.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Surface</th>
<th>Aquifer</th>
<th>Underground</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.66</td>
<td>0.33</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.66</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td>3</td>
<td>0.33</td>
<td>0.66</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.33</td>
<td>0.00</td>
<td>0.66</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td>0.66</td>
<td>0.33</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>0.33</td>
<td>0.66</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Equally distributed weights between the criteria (0.1 for each criterium)

The final map of site suitability was obtained using eq. 8.3:

\[ SI = \sum_{i=1}^{n} W_i \times R_i \]  \hspace{1cm} (8.3)

where \( SI \) stands for suitability index (ranging from 0 - 5, being 5 the most suitable locations), \( W \) and \( R \) correspond to the weights and rating of each criterion \( (i) \), respectively. The site suitability map was subsequently reclassified into 6 classes which accounts for extremely low, very low, low, moderate, high, and very high suitabilities. The different schemas from the sensitivity analysis were also used to create suitability maps in order to compare the different criteria weighting.
The pre-processing, reclassification and rating of all criteria layers and the obtaining of the suitability maps were done through the QGIS software (QGIS Development Team, 2009).

### 8.2.5. Area of interest (AOI) and validation

The selection of an AOI was carried out based on the areas that indicated a high and very high suitability in the map generated. Additionally, low sensitivity to the weighting schemes was also considered for the selection of the AOI.

The validation of the area selected was based on an assessment of the natural recharge in the AOI. In order to evaluate the natural recharge, the master recession curve (MRC) was calculated for different monitoring wells contained in the Australian Groundwater Explorer dataset. From the hydrograph data, bores related to the uppermost aquifer were selected. Additionally, bores whose hydrographs indicated a strong intraseasonal fluctuation were neglected since it was assumed that those were caused by the pumping of irrigation water in the area or a high connection to lower aquifers that are used for extraction. From the MRC, the water table fluctuation method was used to derive recharge from the different hydrographs through eq. 8.4 assuming a limited connection between the different aquifers in the AOI:

\[ R = \frac{\Delta h}{\Delta t} \times S_y \]  

where \( R \) is recharge in a specific time interval \( t \), \( \Delta h \) is the change in groundwater height and \( S_y \) corresponds to the specific yield of the aquifer (Scanlon and Healy, 2002). Since the Namoi River plays an important role in the hydrological and hydrogeological response of the catchment, especially in the alluvial aquifers, recharge was assumed to be a response to rainfall, a losing river system and floods taking place episodically in the floodplain adjacent to the main stream channel. Lamontagne et al. (2015), studying the groundwater geochemistry in the catchment, described two groundwater systems in the alluvial sediments, a shallow one that contains mostly water infiltrated from the river, and a deeper one that contains a mixture of young (river recharge) and old groundwater (GAB). Additionally, Iverach et al. (2017) found the GAB recharge into the alluvial aquifer system as being spatially variable. Since this study concentrates in the shallow aquifer in an AOI defined, and since the GAB discharge into the Namoi alluvium was not distinguishable at most water level monitoring locations, and in particular in the smaller focus area west of Mollee weir (to the west of Narrabri), this input was not taken into account.

Even though the specific yield is a property that changes spatially in the aquifers depending on the composition of the sediments, its estimation is difficult. Therefore,
Chapter 8. Site suitability for managed aquifer recharge

A constant value of 0.05 was used in this study based on Schlumberger Water Services Australia reported value (Schlumberger Water Services Australia, 2012). However, specific yield values between 0.01 and 0.1 are also expected in the alluvial aquifer. Higher values may be found in the main trunk channel associated with a higher composition of sand and gravel particles, while adjacent areas of the floodplain, where the composition of clay and silt particles increases, lead to lower specific yield values.

8.2.6. Water availability

The available water for the MAR project was calculated, constrained to the AOI. It was calculated considering floods as the main source of water for the project.

From the set of hydrological stations in the catchment, a gauge located in highly suitable areas was selected. Its rating curve was evaluated, taking the inflection of the curve as the discharge point at which the stream breaks out of the channel and spreads to the adjacent floodplain. A flood frequency curve was also carried out to evaluate the periodicity of such events and to quantify the available water that might be used for recharge purposes (Kuczera and Franks, 2016).

A schema of the methodology is presented in Figure 8.3.

![Figure 8.3: Flowchart of the methodology used in the study.](image-url)
8.3. Results

8.3.1. Criteria maps

The criteria maps are shown in Figure 8.4. Most of the maps are governed by the geomorphology of the catchment. The slope gradient is one of the main factors creating the landscape and is associated with some hydrogeologic properties. In this case, slope, hydrogeologic units, and groundwater Ks indicate similar patterns.

As it can be seen the most prohibitive criteria, based on the area covered by very low suitability ranks, are those that relate to the distance to rivers and users, and also the aquifer yield, which is mainly constrained by the alluvial deposits in the basin. This is in contrast to the slope map, which shows that most areas in the catchment have low slopes, favorable for MAR projects.

Overall, these maps highlight the importance of stream channels in the catchment development. These, carved in the overlying lithologies due to slope gradients, build up erosional and depositional environments, which finally lead to the generation of
alluvial aquifers and soils. At these locations surrounding the river channels most irrigated agriculture takes places.

8.3.2. AHP analysis

The comparisons between the different criteria layers generate a consolidated decision matrix which can be found in Table A.1 (Appendix section).

Since the pairwise comparison matrix led to a consistency ratio (CR) of 0.013, it was assumed as consistent and we continued with further stages of the decision making process. Subsequently, the decision hierarchy for the evaluated criteria was calculated (Table 8.3).

Table 8. 3. Decision hierarchy of MCDA.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Priority (%)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>18.11</td>
<td>1</td>
</tr>
<tr>
<td>Distance to river</td>
<td>9.26</td>
<td>5</td>
</tr>
<tr>
<td>Distance to users</td>
<td>10.74</td>
<td>4</td>
</tr>
<tr>
<td>Drainage density</td>
<td>8.10</td>
<td>6</td>
</tr>
<tr>
<td>Hydrogeological unit</td>
<td>14.60</td>
<td>3</td>
</tr>
<tr>
<td>Aquifer Ks</td>
<td>5.21</td>
<td>9</td>
</tr>
<tr>
<td>Aquifer yield</td>
<td>4.64</td>
<td>10</td>
</tr>
<tr>
<td>Aquifer salinity</td>
<td>15.26</td>
<td>2</td>
</tr>
<tr>
<td>Groundwater depth</td>
<td>7.69</td>
<td>7</td>
</tr>
<tr>
<td>Soil Ks</td>
<td>6.34</td>
<td>8</td>
</tr>
</tbody>
</table>

As it can be seen, slope, aquifer salinity, and hydrogeologic unit were ranked as the three most important criteria used. Several studies (Ghayoumian et al., 2007; Rahman et al., 2012; Valverde et al., 2016) combine a boolean logic with MCDA, resulting in a first stage in which slope is selected as a criterium to mask suitability areas. This demonstrates the importance of this criterium, especially for surface spreading and in-channel modification MAR methodologies. Additionally, we gave a high importance to the groundwater quality (aquifer salinity) since it would preclude the groundwater use for irrigation purposes, which would make the MAR project useless. The hydrogeologic unit is also highly important since it depicts the medium where the infiltrated water is intended to be stored.

On the contrary, the soil Ks, and the aquifer Ks and yield were considered to have lower importance since these reduce the MAR potential for groundwater recharge,
but would not make it non-viable from the beginning, such as the highly ranked criteria.

### 8.3.3. Site suitability and sensitivity analysis

The suitability maps obtained from the AHP is presented in Figure 8.5. Highly suitable areas for the location of the project are mainly observed close to the stream channels in the alluvial sediments.

![Suitability classes map obtained by coupling the reclassified criteria maps and the AHP technique. High values indicate high suitability.](image)

**Figure 8.5.** Suitability classes map obtained by coupling the reclassified criteria maps and the AHP technique. High values indicate high suitability.

The sensitivity analysis gave similar results in terms of site suitability maps using the different weighting schemes from Table 8.2, except for the schemes 4 and 6 (Figure 8.5), which share a higher relative weighting (0.66) for the underground criteria, including soil Ks and groundwater depth. Therefore, the weighting schemes prioritising surface or aquifer criteria do change the site suitability map results to a lesser extent. This constrains the suitable and highly suitable areas mostly to the
alluvial deposits and whose highly suitable areas can be found in middle sections of the catchment.

**Figure 8. 6.** Suitability maps for the different weighting schemas used in the sensitivity analysis.

An assessment of the areas covered by the different suitability classes is presented in Table 8.4 for the sensitivity analysis. Scheme 3, which assigns a higher relative weighting to the aquifer properties, has the evenest distribution between classes (lower standard deviation), while scheme 7, which uses an even weighting for all criteria, is the most conservative approach for the site suitability map generation with a high deviation and the highest kurtosis. The scheme resulting from the AHP indicates a high deviation between classes, being mostly concentrated between areas of low and moderate suitability.
Table 8.4. Suitability areas for the different weighting schemas used.

<table>
<thead>
<tr>
<th>Schema</th>
<th>Extremely low suitability</th>
<th>Very low suitability</th>
<th>Low suitability</th>
<th>Moderate suitability</th>
<th>High suitability</th>
<th>Very high suitability</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHP</td>
<td>0.00</td>
<td>3.87</td>
<td>53.07</td>
<td>33.78</td>
<td>9.01</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.00</td>
<td>22.97</td>
<td>35.19</td>
<td>32.70</td>
<td>7.58</td>
<td>1.53</td>
</tr>
<tr>
<td>2</td>
<td>0.00</td>
<td>0.04</td>
<td>35.95</td>
<td>56.47</td>
<td>7.37</td>
<td>0.15</td>
</tr>
<tr>
<td>3</td>
<td>0.13</td>
<td>33.44</td>
<td>25.01</td>
<td>28.18</td>
<td>11.04</td>
<td>2.17</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0.19</td>
<td>26.29</td>
<td>49.32</td>
<td>23.96</td>
<td>0.22</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td>2.77</td>
<td>53.74</td>
<td>29.41</td>
<td>13.95</td>
<td>0.10</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>0.20</td>
<td>18.37</td>
<td>54.38</td>
<td>26.59</td>
<td>0.44</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>14.90</td>
<td>57.55</td>
<td>21.24</td>
<td>6.22</td>
<td>0.07</td>
</tr>
</tbody>
</table>

8.3.4. AOI selection and assessment of suitability

Based on the locations that indicate high and very high suitability using the AHP, and with highest agreement across the eight maps of the sensitivity analysis (presented a low sensitivity to the weighting schemas used), an area of interest (AOI) between Narrabri and Wee Waa municipalities was drawn for the location of a MAR project. From this AOI a gauging station with more than 30 years of daily records was selected (Figure 8.7).

**Figure 8.7.** Selection of an area of interest (AOI) for the emplacement of a MAR project, which was located in areas of high and very high suitability. The left image corresponds to the suitability classes’ map, while the zoomed in image in the right corresponds to a false color (bands 4-3-2) Landsat 5 image of the AOI. A gauging
station at Mollee (Number 419039; light green point) and different monitoring wells (yellow points) were selected for the validation and available water assessments.

The selection of the AOI also corresponds to an area where cotton irrigation agriculture is highly developed, which spreads to the west of the catchment. In addition, the overall groundwater gradient of the shallow aquifer follows the slope of the terrain and the stream flow direction, heading toward the west (Mccallum et al., 2009). Several farm dams can be observed in the AOI, which store water for irrigation purposes.

Several monitoring wells (20) are located in the AOI, which present different hydrographic behaviour. Wells located near the Namoi River, in the surrounding paleochannels (abandoned stream channels; Wray, 2009; Welsh et al., 2014), show rapid peaks of groundwater level increase in the hydrograph, which tend to smoothen in wells located farther away (Figure 8.8). These peaks correspond to recharge events (Kelly et al., 2013) and are followed by clear recession behaviour.

Additionally, the AOI selected and all associated monitoring wells are located in areas that according to Iverach et al. (2017) lack of a GAB contribution to recharge or where this is small, and therefore the water fluctuation methodology used for recharge estimation was assumed correct.
Since recharge is a response to natural processes (rainfall, streamflow), it showed to be highly variable in time (Figure 8.9).

**Figure 8.8.** Different hydrographs from monitoring wells located in the AOI.

**Figure 8.9.** Estimation of recharge through the water table fluctuation method in the monitoring well 25107. From a hydrograph (left upper panel) the delineation of...
a master recession curve ($R^2 = 0.97$) was carried out (right upper panel), which leads to a projection of groundwater head changes (left lower panel) and a subsequent estimation of recharges (right lower panel).

Additionally, recharge estimates were also spatially variable (Figure 8.10). Following the trend in the hydrographs, high recharge values were estimated in monitoring wells located in the paleochannel surrounding the Namoi River. Mean recharge estimates range from around 100 to over 700 mm y$^{-1}$. The highest recharge corresponds to a monitoring well located in the western area of the AOI, close to the Namoi River where floods tend to inundate the surrounding area. When compared to mean annual rainfall values obtained from the SILO rainfall dataset, annual recharge estimates in the different monitoring bores range from 2.3 to 21.4% of accumulated rainfall (not shown), with the lower percentages located in the eastern part of the AOI.

**Figure 8.10.** Mean recharge estimates for different labeled monitoring wells located in the AOI.
Since a high spatial variability may be observed in the relatively small area that comprises the AOI, recharge estimates obtained from the water fluctuation method must be interpreted as a local property instead of a regional value (Delin et al., 2007), being highly affected by the underground sediments, which are locally variable, especially in the alluvium in the vicinity of the main stream channel (Kelly et al., 2014).

A correlation between recharge estimates and rainfall was carried out (Figure 8.11 left). In addition, a correlation between mean annual discharge from the selected gauging station and the annual recharge estimates for each bore was also carried out (Figure 8.11 right).

![Correlation coefficients](image)

**Figure 8.11.** Correlation coefficients between recharge estimates and rainfall (left), and between annually aggregated recharge estimates and mean annual discharge at the Namoi River (right).

In general, the mean annual river discharge leads to higher correlations with annual recharge estimates than the disaggregated rainfall. The correlations between discharge and recharge are mostly in the 0.7-0.8 interval, which implies a strong - very strong relationship between variables. In contrast correlation coefficients between rainfall and recharge estimates are mainly in the 0.4-0.7 interval, indicating a moderate - strong relationship between variables.

The high recharge estimates in the AOI, which are higher in the vicinity of the Namoi River, indicate locations where the natural recharge can be used for a MAR project. Since recharge events in the study area are strongly dependent on the streamflow, spreading methods can be implemented in the palaeochannel surrounding the Namoi
Chapter 8. Site suitability for managed aquifer recharge

River, but also in-channel modifications can be developed to increase the recharge from the Namoi River.

### 8.3.5. Water availability from gauging data

The hydrograph analysis of the selected gauging station is in Figure 8.12. It presents the rating and flood frequency curves of the Namoi River at Mollee. Additionally, a logarithmic regression was fitted to the flood frequency data (dashed line).

![Figure 8.12](image)

**Figure 8.12.** Rating (left) and flood frequency (right) curves at the Mollee gauging station selected. The flood frequency curve presents both, a logarithmic scale in both axis and a logarithmic regression curve (dashed line).

From the hydrograph analysis, a discharge value of approximately 500 m$^3$ s$^{-1}$ leads to an inflection in the rating curve, which was assumed to be the discharge that leads to a flow that exceeds the channel storage capacity, and therefore leads to floods over the surrounding floodplain. Then, assuming that the MAR project will capture at least one flood event within a return period of 5 years to replenish the groundwater storage, this would correspond to a daily mean discharge of 883 m$^3$ s$^{-1}$. Subtracting the discharge threshold obtained from the rating curve to this discharge at a 5 years return period leads to 383 m$^3$ s$^{-1}$ exceeding the stream channel capacity and causing flooding in the surrounding areas, which implies a volume of over 33 GL per day that might be allocated for recharge purposes. This volume constitutes around 8% of the Keepit dam capacity, the major dam in the catchment. However, return period were estimated using the annual maximum
series, while actually flood waves might take several days in traveling and receding (Fuentes et al., 2019). This implies that floods usually take longer than a single day, and therefore greater available water volumes for recharge are expected from such events. For instance, the flood occurred in November 2011 in the Molle gauging station took around six days in receding, with discharges over 800 m$^3$ s$^{-1}$ (5 consecutive days over 1000 m$^3$ s$^{-1}$), which implied a volume of over 240 GL exceeding the channel, more than half the storage capacity of the Keepit dam, which might be used to enhance recharge.

8.4. Discussion

There are many studies coupling MCDA and GIS to evaluate the site suitability for MAR projects (Rahman et al., 2012; Russo et al., 2014; Kazakis, 2018; Sallwey et al., 2018). However, several lack sensitivity analysis or validation of results (Ghayoumian et al., 2007; Chowdhury et al., 2009; Rahman et al., 2012; Owusu et al., 2017). In this study, the sensitivity analysis not only showed the effect of changing the criteria weights on the site suitability, but it also allowed the selection of suitable areas, which were less sensitive to the weighting decision.

In this study, the weight assigned to the permeability of the soils criterium was ranked the lowest, in opposition to Arshad et al. (2014). This is explained by the fact that the current study assessed mainly the physical suitability of MAR projects rather than its economic feasibility. Since the permeability of the soil interface can be either enhanced or bypassed by using different artificial recharge techniques (Bouwer, 2002), it becomes less significant than other criteria from a physical perspective.

Validation is one of the most variable aspects in site suitability for MAR projects. For instance, Saidi et al. (2017) suggest that there is no direct method to verify MAR suitable sites, and therefore indirect methods must be applied. Some studies (Ghayoumian et al., 2007; Mahmoud, 2014) have used the presence of ongoing MAR projects to validate their results. Kazakis (2018) assumed the consistency of the AHP technique and a subsequent sensitivity analysis to be enough for validation, while Russo at al. (2014) used a numerical model to validate their results. In the present study, the natural recharge estimated through the water table fluctuation method in different monitoring wells was used to validate the AOI delineated since the preferred methods of recharge imply either spreading alternatives or in-channel modifications.

One of the difficulties of estimating recharge is its spatio-temporal variability, which is frequently neglected (Manna et al., 2019). It is also difficult to get recharge estimates due to the limitations of the different methodologies available for its quantification. Knowing this difficulty, Scanlon et al. (2002) proposed the use of
combined alternatives for a recharge estimation, which makes the estimation harder.

There are several advantages and disadvantages in relation to the water table fluctuation methodology. While it allows an estimate of groundwater recharge in time, it depends on other properties, such as the specific yield (Healy and Cook, 2002), which was assumed constant in this study. Additionally, the connectivity between aquifers was neglected based on the hydrograph behavior. However, Lamontagne et al. (2015) and Iverach et al. (2017) studies present enough evidence to validate this approach used. The temporal frequency of the recorded water levels tended to vary in the dataset. During some of the time intervals sampling happened at low frequency (such as quarterly), which might affect the recharge estimates. Nevertheless, since the hydrograph data is publicly available and has long time series, it constitutes a valuable data source for local recharge estimations.

Different studies have estimated recharge in the Namoi catchment with quite variable results depending on the methodologies and study locations. Crosbie et al. (2010) identified a recharge of 58 mm y$^{-1}$ in the lower Namoi area using the water table fluctuation. McCallum et al. (2010) by using the WAVES model estimated 14 mm y$^{-1}$ of recharge, but this was only valid for the specific date and location studied, and assuming recharge was equivalent to drainage below 4 m depth. Abbs and Littleboy (1998) used the PERFECT cropping model to evaluate recharge in the Liverpool Plains (upper Namoi catchment), and estimated between 28 and 80 mm y$^{-1}$ in agricultural black earth soils. Weaver et al. (2013) using a chloride mass balance found deep drainage estimates from a few mm to around 400 mm between the 2010-2011 cotton growing season at three sites around Namoi, Wee Waa and Merah North, all located in the lower Namoi. Cuthbert et al. (2016), by studying an ephemeral stream tributary of Maules Creek in the upper Namoi, estimated from 30 to 80 mm y$^{-1}$ of indirect recharge by using the water table fluctuation method. Timms et al. (2012), studying two sites located in floodplains of the lower Namoi, modelled deep drainage using APSIM and estimated drainage values between 3.3 and 9.5 mm y$^{-1}$. All these studies prove that groundwater recharge in a catchment is spatially variable.

Coincidently in the selected AOI, Kelly et al. (2009) evaluated the Namoi River between the Mollee (close to Narrabri) and the Gunidgera (near Wee Waa) weirs using continuous electrical imaging for mapping recharge, and found potential recharge pathways in this reach of the Namoi River. The same study describes that the hydraulic connectivity between aquifers increases to the west, beyond the AOI selected in this study, and that in the range of longitudes that coincide with the current AOI is where most of the recharge occurs due to the presence of sand and gravels in the upper 10 m of sediments underlying the river, which is also supported by some cross sectional profiles in Lamontagne et al. (2015; Figure 8.13). These
characteristics help to explain the variability and the high recharge estimates in the AOI, which are also reported in Williams et al. (1969) for the Gunidgera and Mollee weir locations.

![Cross section profile of the hydrogeological setting derived from boreholes at Mollee.](image)

**Figure 8. 13.** Cross section profile of the hydrogeological setting derived from boreholes at Mollee.

Lamontagne et al. (2015) studying the river infiltration in a reach of the Namoi River between the Molle and Yarral weirs (both included in the AOI), found that the river behaved as a losing system most of the time, except during floods, in which the river led to bank recharge-discharge cycles. Using hydrogeochemical analysis, the study estimated infiltration rates ranging from 0.6 to 5 m yr⁻¹. Iverach et al. (2017), studying the groundwater mixing process in the Lower Namoi alluvium, also found significant recharge during floods, especially in monitoring wells near the river, which present a high connectivity with the surface.

This study estimated high recharge values, which are nevertheless within the range of recharge values found in the aforementioned literature. These estimates are justified based on the locations where they were obtained, which correspond to the surroundings of the Namoi River, in relatively high conductive sediments. Therefore, these values may be used as a way to validate the results of the suitability map.
A previous study of the potential for a MAR project in the catchment was carried out by Woolley et al. (1994). This identified potential sites for a MAR trial located in some abandoned gravel pits near Merah north in the lower Namoi catchment. The recharge in the area is described as small except when overbank flooding occurs. The study also suggested potential trial sites in Doreen Lane, located 9 km to the north of Merah north, but without specifying how to transport the water for the trial project. Two other sites were defined in the upper Namoi catchment, close to Quirindi through the use of a weir, and in the Watermark area. Most of the areas mentioned were classified as having high suitabilities in the current study, but were not as insensitive as the AOI selected. Woolley et al. (1994) also estimated a monthly availability of 30 GL of water for a MAR project. However, their study suggests to take such estimate with extreme caution, and that it would coincide with floods.

Since weirs increase the opportunity for river infiltration and the hydraulic head, it is clear that these present a high potential for inducing recharge in the river (Kelly et al., 2013), especially if these are located in highly conductive reaches. Hydrograph data gives clue of such unintentional benefit in the Mollee weir. However, induced recharge estimates in this location have not been carried out. These structures might be one of the ways to conceive the development of a MAR project in the catchment, taking advantage of the high conductivity of underlying sediments. Nevertheless, groundwater monitoring bores are still too sparse and spaced to identify the higher conductive locations in the paleochannels surrounding the Namoi River. Therefore, a higher spatial resolution to find optimal recharge locations by geophysical surveys and detailed lithological coring in the region is still required for the emplacement of a MAR project that may operate under spreading or in-channel modification techniques.

There are several studies that discuss the role of floods in groundwater recharge. Doble et al. (2013) in another Australian catchment found that floods contributed a 4% of the total recharge, being 17% of the riparian recharge. Kelly et al. (2013) discussed the peaks in Namoi groundwater hydrographs as depicting flood recharge. Regarding the river system, Brownbill et al. (2011) described the studied reach of the Namoi River as a connected losing system, which is recharging groundwater. Since during floods the water levels in the stream channels also rise, these lead to a higher hydraulic gradient, which implies higher recharge, also explaining the dependence of recharge on the river discharge.

Although available water estimates from floods can be calculated, that does not necessarily mean that such volumes can be used for irrigation. Regulations and entitlements are strong constraints on the potential use of available water even in surplus periods, such as during floods. Since MAR has got a relatively slow acceptance (Dillon et al., 2018), regulations regarding its applicability are still a
limiting factor. However, increased recharge into the groundwater can have additional benefits in relation to overall groundwater sustainability. Several MAR projects that use spreading or in-channel modification systems are in operation, however these are considerably less popular than bank filtration (Sprenger et al., 2017), which is mainly used as a water treatment system (Maeng, 2010). Spreading methods to enhance recharge have historically been used, and are currently the second most common technique to implement MAR in Europe, with 77 active sites in 2013 (Sprenger et al., 2017). However, such projects have also been implemented successfully in other continents. For instance, the development of pilot projects for MAR under spreading methods has been implemented since the 1960’s in Arizona (United States; Dillon et al., 2018), and are envisaged in the present as important ways to recharge depleted aquifers in the Central Valley of California by on-farm recharge flooding (Kocis and Dahlke, 2017). In Japan, a spreading basin operating since 2009 in alluvial fans has been reported as contributing to the sustainable management of the local aquifer (Hida, 2009). Regarding in-channel modification projects, Sprenger et al. (2017) only mention one active project in Europe, while Dillon et al. (2018) describes that it has been applied extensively in India, but also in China.

Even though the site suitability map generated indicates potential areas for a MAR project implementation, and even though a validation of one of the potential sites was carried out, the numerical modeling of different artificial recharge alternatives should be carried out to assess the impact of such project on the surroundings. Additionally, no water quality assessment has been considered in this study, which is also required before the implementation of a pilot project. Therefore, further studies addressing these issues and analysing other MCDA alternatives and their corresponding uncertainties will be matter for future investigations. Finally, the modeling of the groundwater response to the MAR project located in the selected AOI might also be a subject for future research.

8.5. Conclusions

The combined use of MCDA and GIS is a valuable tool for the selection of suitable sites for MAR projects. These can allow a consistency analysis and represent useful alternatives that can shed light on territorial planning through the use of publicly available data in low local data environments. This resulted finally in the generation of a site suitability map. Additionally, sensitivity analysis allows an assessment of criteria weighting and the selection of AOIs to locate potential MAR projects at a catchment scale. However, the selected MCDA does not allow for a quantification of the uncertainties in the weighting process, and strongly relies on the user’s judgment.
to prioritise different criteria, having restricted applicability for more automatised tasks.

The validation process used estimates the natural groundwater recharge to highlight its spatial variability, but also showed that natural recharge was strongly dependent on the river streamflow patterns. This analysis also confirms how floods and streamflow are a significant source of recharge in the catchment. Moreover, recharge estimates in the selected AOI are high, yet higher estimates can be associated to the paleochannel surrounding the Namoi River. Among the different techniques for MAR projects, weirs might be built in conductive reaches of the stream network to enhance groundwater recharge.

This study provides potential locations and spatio-temporal estimates of locally available water for MAR, which might be derived from floods by using spread or in-channel modification methods. However, the practical use of this water for MAR and irrigation would be limited by local regulations and water entitlements. Even though this study finds potential sites for a MAR project, geophysical surveys or detailed lithological coring might further constrain the selection of optimal sites based on the conductive properties of the sediments.

8.6. Appendices

Table 8. A 1. Consolidated pairwise comparison matrix obtained through the AHP technique. Values close to 1 mean variables have equal importance, values smaller or greater than 1 indicate smaller or greater importance of the row criteria compared to the column criteria.

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1.00</td>
<td>2.45</td>
<td>2.45</td>
<td>1.73</td>
<td>1.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.41</td>
<td>2.83</td>
<td>2.45</td>
</tr>
<tr>
<td>C2</td>
<td>0.41</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.71</td>
<td>2.45</td>
<td>2.45</td>
<td>0.58</td>
<td>1.00</td>
<td>1.22</td>
</tr>
<tr>
<td>C3</td>
<td>0.41</td>
<td>1.00</td>
<td>1.00</td>
<td>2.00</td>
<td>1.00</td>
<td>1.41</td>
<td>2.00</td>
<td>0.71</td>
<td>1.41</td>
<td>2.00</td>
</tr>
<tr>
<td>C4</td>
<td>0.58</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
<td>0.50</td>
<td>1.73</td>
<td>1.73</td>
<td>0.71</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>C5</td>
<td>1.00</td>
<td>1.41</td>
<td>1.00</td>
<td>2.00</td>
<td>1.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>2.00</td>
<td>2.45</td>
</tr>
<tr>
<td>C6</td>
<td>0.33</td>
<td>0.41</td>
<td>0.71</td>
<td>0.58</td>
<td>0.33</td>
<td>1.00</td>
<td>1.00</td>
<td>0.29</td>
<td>0.71</td>
<td>1.00</td>
</tr>
<tr>
<td>C7</td>
<td>0.33</td>
<td>0.41</td>
<td>0.50</td>
<td>0.58</td>
<td>0.33</td>
<td>1.00</td>
<td>1.00</td>
<td>0.33</td>
<td>0.50</td>
<td>0.58</td>
</tr>
<tr>
<td>C8</td>
<td>0.71</td>
<td>1.73</td>
<td>1.41</td>
<td>1.41</td>
<td>1.00</td>
<td>3.46</td>
<td>3.00</td>
<td>1.00</td>
<td>2.45</td>
<td>3.00</td>
</tr>
<tr>
<td>C9</td>
<td>0.35</td>
<td>1.00</td>
<td>0.71</td>
<td>1.00</td>
<td>0.50</td>
<td>1.41</td>
<td>2.00</td>
<td>0.41</td>
<td>1.00</td>
<td>1.41</td>
</tr>
<tr>
<td>C10</td>
<td>0.41</td>
<td>0.82</td>
<td>0.50</td>
<td>1.00</td>
<td>0.41</td>
<td>1.00</td>
<td>1.73</td>
<td>0.33</td>
<td>0.71</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* C1: slope; C2: distance to river; C3: distance to users; C4: drainage density; C5: surface hydrogeological units; C6: aquifer Ks; C7: aquifer yield; C8: aquifer salinity; C9: groundwater depth; C10: soil Ks
8.7. References


Chapter 8. Site suitability for managed aquifer recharge


Chapter 8. Site suitability for managed aquifer recharge


Chapter 8. Site suitability for managed aquifer recharge


Chapter 8. Site suitability for managed aquifer recharge


Chapter 8. Site suitability for managed aquifer recharge


Chapter 9

General discussion, conclusions, and future work

9.1. General discussion

Among water management alternatives to increase water security, the building of dams has been probably the most commonly applied (Steduto et al., 2012). However, evaporation losses can be substantial for reservoir storage (Craig, et al., 2005; Zhao and Gao, 2019). The second chapter of this thesis indicates that the development of a MAR project might have a beneficial impact for water usage. Since open water evaporation per surface water unit is significantly increasing in the Namoi basin, strategies to avoid evaporation such as MAR (Casanova et al., 2016; Scanlon et al., 2016) are fully justified. Additionally, because reference evapotranspiration and rainfall are increasing and decreasing respectively, water management practices should match the pressures that are imposed by climate change and enhance water security (Turral et al., 2011).

However, even though MAR seems fully justified, the water source for such a project is still a matter of debate. Even if floods volumes have the potential to replenish aquifers, these large volumes are also affected by climate variability (Kiem et al., 2003). In this study, changes in flood volumes were not assessed. Bown and Rivera (2007), studying climate change effects in Chile, pointed out an increase in the frequency of droughts caused by the El Niño Southern Oscillation (ENSO). Interannual climate variability modulated by ENSO also strongly affects rainfall patterns in Australia (Nicholls, 1991). Therefore, studying flood variability, in terms of frequency and magnitude, under climate change may lead to a better understanding of water availability (Paltan et al., 2017). Hirabayashi et al. (2013), studying global flood changes due to climate change, describe spatial variability of flood frequency across the different continents. For the Murray-Darling basin the authors predict an increase of frequency in floods for the XXI century. A similar result was obtained by Milly et al. (2002) using models with increased CO₂ concentrations. However, it would be better to use the available discharge data, to evaluate actual temporal changes.

As pointed out in Chapter 2, impacts of climate change and water usage effects on water availability still need to be disentangled. For instance, it is not clear if surface water changes are a consequence of climate or water use, or if surface water changes are a combination of these, and the contribution of each one of them on such changes is not clear. The same can be suggested for groundwater resources. It is not clear if groundwater levels decline, recorded in some wells (Kelly et al., 2014), are a response to improper water management and overexploitation, or if it
is the result of lower recharge associated to a decrease in rainfall, land use changes, or a combination of these.

In this thesis, different factors related to the potential for a MAR project implementation were studied as shown in Figure 9.1. The different products might be used as inputs for the development of a groundwater model to evaluate the potential impacts of the MAR project. Finally, this can be empirically evaluated through pilot projects.

**Figure 9.1.** Main schema summarising the different products obtained in this thesis, their methodologies and how these connect with future works for the development of a MAR project.

### 9.1.1. Remote sensing contribution

As shown extensively in this thesis, remote sensing has increasingly allowed monitoring of surface water resources. However, hydrologic systems, which can be roughly divided into surface water and groundwater, must be studied jointly in order to understand the different water budget components. These allow estimation of water availability in different regions, which is increasingly needed under the uncertain scenarios caused by climate change (Dessai and Hulme, 2007). In terms of available water for MAR projects, two different methodologies were applied. In Chapter 4 flood volumes derived from remote sensing were estimated, while in Chapter 7, an estimation was carried out using gauging station data.

Both methods used a frequency analysis of the data, and led to differences of around one order of magnitude. In the case of floods monitored by gauging stations, the
maximum annual peaks were used for the flood frequency curves (Kuczera and Franks, 2016). However, as stated in chapter 8, floods may take several days in receding, depending on the catchment size, topography, land cover, among others. Therefore, flood volumes may be the accumulation of several days of discharge exceeding stream channel capacities, such as occurred in 2012. Additionally, stream networks often contain several crevasse splays, i.e. locations in which the streamflow breaks into the floodplain (Bristow et al., 1999), which affects the estimates of flood volumes at gauges. In contrast, remote sensing volume estimates imply total accumulated water causing floods. Even if the frequency analysis of images is carried out using maximum annual peaks, these imply a snapshot of streamflow heights, and can only be obtained after clouds vanish (Schumann and Moller, 2015), which implies some infiltration and evaporation losses have taken place.

High resolution remote sensing products can lead to more accurate results in terms of water availability using remote sensing (Feng et al., 2016; Pekel et al., 2016), but result in high computational requirements when evaluated at large scales. At this stage of computational development, memory limitations still constrain the analysis and handling of big data. However, the spatial resolution is not always the main limiting factor for studying hydrologic processes. As pointed out in the third Chapter, temporal resolution for some processes is as significant as the spatial resolution (Khan et al., 2011). Since floods correspond to anomalies in time series of climatic conditions, the ability to capture their magnitude and extent strongly relies on the frequency to monitor the processes in time. Thus, MODIS products with a daily frequency are more useful for flood forecasting than other products with higher spatial resolution, such as Landsat or Sentinel imageries.

For the long-term monitoring of processes, the longer the monitoring period the better. That is the reason why in Chapter 2 and 4 Landsat imageries were used rather than MODIS or Sentinel (Pekel et al., 2016). By combining the images of the Landsat constellation it is possible to get over 30 years of data, which covers the period included for climate analysis (30 years). However, due to its sparse temporal resolution, these are more useful to study average conditions, or processes taking place at slow rates. Since water trends were evaluated in Chapter 2 and reservoir volume changes in Chapter 4, Landsat images were found to be convenient for such analyses.

The Surface Water and Ocean Topography (SWOT) mission, whose launch is projected in 2021 with the objective to monitor ocean and inland waters, might improve surface water availability monitoring. However, since other satellites have accumulated several years of data, their products are still valuable information to study relatively long-term processes and changes.
9.1.2. Geologic framework

Different sources of data can be used to analyse Earth systems. In this case, in Chapter 6 and 7 it is demonstrated that basic textual data, but with a corresponding geolocation, can be used for lithological mapping and modeling. The use of NLP for environmental sciences has rarely been explored until now, and with the development of new algorithms and techniques, can facilitate the processing and analysis of data that has not been used due to the numerical bias of sciences. This implies new opportunities for the mapping and modeling of geologic materials and soils, which have been extensively surveyed and described (Culshaw, 2005), and can be used as a tool for geology mapping, soil taxonomy and classification in conjunction with machine learning.

Even with three dimensional models of the lithologic setting, it is not totally clear if these could imply a benefit for groundwater modelling. The use of well-defined geological schemes in groundwater modeling has been frequently oversimplified (Dickson et al., 2014). This is mainly because of a lack of geological information that hampers the depiction of aquifer geometry and groundwater flow in three dimensions (Turner et al., 2015). However, it must be considered that models correspond to a simplistic representation of reality to make realistic predictions (Anderson et al., 2015). Therefore, a complex three dimensional representation of geology, apart from increasing memory consumption for the model, might not necessarily lead to any practical benefit, and therefore should be investigated.

In this regard, while the classic hydrogeological setting in the Namoi catchment has divided the alluvial aquifer in three aquifer units (Williams et al., 1969; Welsh et al., 2014), recent studies have criticised this conceptualisation mainly based on the hydrochemistry of groundwater (Iverach et al., 2017; Lamontagne et al., 2015). Based on these studies, the shallow aquifer corresponds to a distributive fluvial system with rainfall, river and GAB aquifer contributions to recharge. Based on this, a main channel upstream in the catchment divides into several smaller meandering channels, and the hydraulic conductivity of these depend on their sediment composition (Kelly et al., 2014). Additionally, surface recharge locations would be based on the composition of sediments, and enhanced in areas where sandbars occur and have occurred in current stream channels and ancient fluvial systems, respectively, that connect with the shallow aquifer.

A lithological 3D map for the area in Chapter 8 is shown in Figure 9.2. Clearly, the distribution of sediments does not represent continuous layers in depth, but a heterogeneous distribution in the region, with different degrees of continuity in depth, which might support the claim that the alluvial aquifer represents a distributive fluvial palaeo-system. Coarse sediments outcrop in some central areas, close to the stream channel, but they can also be found at shallow depths in southern.
areas. Additionally, volcanic outcrops can be found to the northeast, which agrees with the hydrogeological setting mapped by the groundwater geofabric dataset.

![Lithological 3D model of the AOI selected in Chapter 8 in the Namoi catchment obtained using a triangulation of embeddings.](image)

**Figure 9. 2.** Lithological 3D model of the AOI selected in Chapter 8 in the Namoi catchment obtained using a triangulation of embeddings.

### 9.1.3. Groundwater models

The MCDA used in Chapter 8 is a useful alternative to identify areas suitable for MAR projects. However, the feasibility should be ultimately tested through the development of pilot projects to empirically evaluate the changes in water storage and the potential impact on aquifer properties, water quality, and water dependent ecosystems. Numerical models may precede pilot projects to evaluate the response of groundwater to recharge volumes. These also have to be considered as theoretical evaluations and subsequently empirically tested. This could be built on the different regional models that have been developed for the different areas in the Namoi catchment.

In this thesis, floods imply a significant source of water for groundwater recharge. However, Rassam et al. (2008) describe groundwater models in the Murray Darling basin as deficient in terms of considering floods as a source of recharge. By including the surface water-groundwater interaction in a numerical model of the Namoi river, better low flow groundwater predictions and also a more realistic calibration of the model were achieved (Rassam et al. 2013).
Chapter 9. General discussion

Several gaps in the groundwater knowledge of the study catchment have been identified. Kelly et al. (2009) present a description of the Upper and Lower Namoi groundwater models that have been developed. These consider the classic division into aquifer units rather than assuming these as a single distributive fluvial system in the discretization stage. Other gaps mentioned are deficiencies in taking into account groundwater-surface water interactions and deep drainage, and identify the need of mapping zones of recharge and discharge along the river, and the potential for managed aquifer recharge projects. As demonstrated, MCDA can be used as a prospective tool to constrain areas of recharge and to detect locations for a MAR project. In the case of the Lower Namoi groundwater model, recharge has been evaluated as relying strongly on episodic floods; however, in the long-term, the greater source of recharge corresponds to stream losses, contributing on average to 41 GL y⁻¹ between 1980 and 1998 (Kelly et al., 2009). This main component of river losses on groundwater recharge in the catchment should be taken into account when making decisions regarding the type of MAR project to implement.

Groundwater extraction is a difficult groundwater component to estimate. In the Upper and Lower Namoi groundwater models described by Kelly et al. (2009), pumping was considered quite simplistically. The current Upper Namoi model assumes an average of 58 GL y⁻¹ of pumping. In the case of the Lower Namoi, groundwater extraction is variable (Welsh et al., 2014). According to Smithson (2009) groundwater usage between 2006 and 2007 in the Lower Namoi was of 125.7 GL, yet it decreased to 102.4 GL in 2007-2008. However, since groundwater is mainly used for irrigation, its occurrence is strongly seasonal. One of the major concerns related to heavily pumping freshwater from deep alluvial aquifers in the catchment is the potential migration of shallow saline waters towards the highly productive freshwater aquifers, which requires more attention (Kelly et al., 2009). Likewise, when considering a MAR project implementation, its location must be selected where no risk of saline contamination exist.

One of the conclusions of this thesis is that the transmissive properties of the infiltration medium, which include recharge, are spatially variable in the catchment depending on the sediment composition. Therefore, extrapolating recharge values for different management areas in the catchment might also affect model predictions, especially when considered at finer scale resolutions. Similarly, assuming a single hydraulic conductivity for the different aquifer units is an unrealistic approach, but this is widely used for modelling purposes, and translates into prediction errors.

9.1.4. General MAR characteristics

One of the main aspects considered in this study was the use of surface water derived from floods for the development of a MAR project. This was decided in order
to avoid the development of an aquifer storage recovery project due to the increase in energy consumption for water injection, and because of the requirement of water treatment. These significantly increase the costs of the project and according to Arshad et al. (2014) might make it infeasible.

However, the quality of flood waters should also be studied to avoid aquifer contamination (Casanova et al., 2016b), even though the infiltration and transport through soils and sediments might improve water quality by filtering and adsorption of pollutants (Sharma and Kennedy, 2017). Additionally, it must be considered that all projects involve clogging. Clogging takes place at the surface where water is infiltrated/injected and through the transport medium. It occurs through time and reduces the transmissivity of the transport medium (Bekele et al., 2015). While clogging in injection wells takes place through the sieve and walls of wells and in the aquifer porosity, which can be reduced knowing the chemistry of water, in spreading methods clogging mostly occurs through the deposition of fine sediments in the surface. Clogging in this case, might be reduced through decantation pools, or the infiltration basins can be periodically treated by removing deposited sediments at the surface (Martin, 2013). However, in spreading methods clogging might also take place more slowly by filling the pore system of the exchange medium. This clogging rate will ultimately define the projection of the operation time for the project.

While MAR can contribute to the water storage capacity in a catchment, it would also alter the natural functioning of the system, which might decrease water runoff and floods. This can have important effects for ecosystems that rely on periodic floods and riparian ecosystems (Dillon et al., 2009). A similar example of water management interventions affecting ecosystems can be found in Overton and Doody (2008) that explains how the Hume dam and the Australian Murray river regulation altered the flood dynamic of the catchment, endangering flood dependent ecosystems. Therefore, studies to evaluate the environmental impact of such systems are also required. In the end, adequate water management must offset different potential benefits and disadvantages in order to preserve socio-economic development and ecosystem functioning.

### 9.1.5. Regulations and economic aspects related to MAR

This thesis is mainly focused on physical aspects for the location of MAR projects. Therefore, regulations and economic feasibility were briefly commented on, but neglected in most analysis. For instance, the Namoi Water Resource Plan divides the catchment into regulated, below main reservoirs, and unregulated systems, above main reservoirs, and for these, a division exists between Upper and Lower Namoi and Peel valley. Additionally, the available water in the catchment is used to supply water to different licences, which include basic rights, environmental water, general security licences, high security licences, and supplementary licences. Water
resources management has been agreed as part of the Upper and Lower Namoi regulated systems in the Water Sharing Plan for the Upper Namoi and Lower Namoi Regulated River Water Sources 2016, which started on 2016 up to 2026, in compliance with the Water Management Act 2000 (Burrell et al., 2018). On the other hand, the groundwater resource management is specified separately in the Water Sharing Plan for the Upper and Lower Namoi Groundwater Sources 2019, which divides the Upper Namoi in 12 groundwater zones, and the Lower Namoi into a single zone. These management divisions, which consider the surface waters and groundwaters as separated systems, might be considered as deficient due to the strong interaction between both systems.

Any management implemented in the catchment should follow the directions of the water sharing plan in terms of surface water. Therefore, a MAR project should be based on this agreement. The Upper Namoi in the period 2017-2018 had 11,454 shares for general security licences, and 685 shares for other licence types. For general licences, shares correspond to 1 ML. Furthermore, in terms of usage, 8% of the general security share component was inactive, and the use of available water from regulated sources was 89%. In the case of the Lower Namoi, the general security component accounted for 245,315 shares, and 115,480 shares of supplementary water licences. However, no water was obtained through supplementary events since the threshold in the water storage was not reached. Usage accounted for 175 GL for the 2017-2018 period, and 4% of the general security share component was inactive. In conjunction, both the Upper and the Lower Namoi results in a usage of 180 GL, with a moving average in the last years of 132 GL (Burrell et al., 2018). Additionally, there is an environmental share component for both the Upper and Lower Namoi, and stakeholders may reserve shares for the subsequent year (carryover).

Thus, even though floods imply surplus water, these volumes are allocated in the supplementary water share component. Therefore, its use is restricted to shareholders. In addition, since MAR is mainly intended as a storage water alternative, it implies subsequent pumping of the water for usage (irrigation). This precludes any environmental water usage for such a project. As a consequence, water for a MAR project should be considered within the licence framework, which limits the application. For a proper MAR management under the current water sharing plan, collaboration among supplementary water stakeholders seems as one of the best alternatives for implementation. However, under the current regulations, MAR projects still have strong uncertainty for stakeholders, which makes its application even more difficult.

An important aspect of the feasibility of MAR projects such as the one discussed here, is the agricultural focus. While several MAR projects are under operation associated to mining and water supply activities, agricultural projects rarely integrate MAR as one of the water sources. Since agriculture projects are not as profitable as mining projects, the economic rationale for its
implementation/operation is not as clear. This further constrains potential MAR alternatives. As pointed out by Arshad et al. (2014), aquifer storage and recharge projects in the Namoi catchment are unfeasible under current water conditions (water prices in the market), and low infiltration rates of fine surface sediments in Namoi may make projects infeasible. This is one of the reasons that justifies a more detailed geophysical survey to find optimal locations of high recharge where to locate a preliminary project. Dillon and Arshad (2016) evaluated the feasibility of MAR projects in the framework of integrated water resources management, and concluded that infiltration basins with moderate-high hydraulic conductivities imply the highest net benefits. In this case, and due to the difficulty to find such locations with moderate-high infiltration rates, in-channel modifications could be used for the same purpose.

However, even if these locations for recharge can be found, and if the water budget allows available water for the project and favourable regulations exist for implementation, there is still a need to evaluate the acceptability of MAR in the surrounding communities and among stakeholders. Earlier, Rawluk et al. (2013) evaluated the acceptability of large flood waters for MAR projects in the Namoi valley. This showed that even though MAR is recognised as having some benefits, it also causes concern due to its potential impact on aquifers, but also because it could cause over-exploitation of water resources. Given these reasons, it is clear that further science must address these concerns.

9.2. Conclusions

Climate trends for the 1988-2018 period showed an intensification in the hydrologic cycle in the Namoi catchment, with increasing temperatures and reference evapotranspiration over a large area, but also with decreasing rainfall and humidity in some areas. These imply an overall reduction of available blue water for usage. Even though on average surface water bodies did not indicate a trend, these were spatially variable. Strong evidence for negative trends in open water frequency was observed in natural reservoirs, such as the Goran lake, while positive trends were found in several farm dams. In general, a strong positive trend in open water evaporation per unit surface water was detected, which justifies water management alternatives such as MAR for storing water while reducing evaporation losses.

Remote sensing surface reflectance data characterised the spatiotemporal behaviour of floods at the catchment scale, firstly in terms of flood extent, but subsequently, combining surface reflectance data with a DTM, for flood volume estimates. Here, a high temporal resolution of images was favourable to capture flood extents. The frequency of flood volumes was also obtained, which highlighted the possibility of using runoff water to enhance infiltration and recharge aquifers.
Using NLP techniques, such as building word embedding models, gave a fair representation of geologic texts in a multidimensional space which could capture the relationship between associated concepts. Using this, in conjunction with a fully connected neural network, could accurately classify borehole descriptions into lithological classes. Additionally, by applying simple interpolation techniques to the embeddings, 3D lithological maps were built which gave a reasonable representation of the hydrogeologic settings. These allow a better understanding of the lithology associated with groundwater transport.

The combination of MCDA with GIS demonstrated to be a useful alternative to mix different sources of information (criteria) in the development of site suitability maps for the location of MAR projects. Additionally, using a sensitivity analysis and the site suitability map allowed the selection of an area, to further constrain the search of a location for the project location. Groundwater recharge, estimated through hydrograph analysis, confirmed areas of high infiltration associated with the paleochannel of the Namoi river, which coincided with the shallow presence of coarse sediment deposits. Since this area is close to the Namoi river, techniques using inchannel modification are suggested as potential MAR technique. Among these, weirs in the stream channel currently induce unintended recharge, which might be strengthened if recharge needs to be increased.

9.3. Future work

This study includes several topics that require further studies. In terms of climate, since floods are episodic anomalies in rainfall patterns, and since these are affected by interannual climate variability, the study of flood variability and change is required. This includes floods and droughts, and the change in temporal patterns, including frequency and intensity. Additionally, flood forecasting requires more attention, given predicted increasing flood frequencies caused by climate change (Hirabayashi et al., 2013).

A simple pooling alternative was used in this study to ensemble different flood algorithms by using a linear equation with equal weights. But other alternatives have been proposed, which have led to better performances. Combination of probability results through different methods might help to obtain more realistic representation of studied processes.

As mentioned in the general discussion, since climate and water usage have a combined effect on the change in water resources, it becomes important to disentangle both signals in order to understand their individual contributions on the depletion of water resources. This can contribute to water management plans at regional scales.
In relation to the lithologic modeling and mapping using NLP and machine learning, two main deficiencies were identified. Firstly, the classification was supervised and therefore this implies a lack of automation in the 3D lithological mapping. As a consequence, different unsupervised algorithms should be tested to fully automate the build up of lithologic maps. Secondly, interpolation techniques in the three dimensional space showed several deficiencies in relation to the accurate prediction of the lithologies. Thus, the spatial distribution of borehole descriptions should be considered in the model. These are widely spaced in longitude/latitude axes, but closely spaced along the elevation axis. Therefore, the performance in 3D lithologic predictions might be improved by using machine learning algorithms.

While the site selected for the MAR emplacement using MCDA was consistent with recharge estimates of boreholes, the technique lacked an uncertainty analysis. However, uncertainties during the MCDA steps are inherent to the methodology since arbitrary decisions are made by users. Therefore, an assessment of uncertainty may be addressed by combining the MCDA with simulation algorithms.

The evaluation of a MAR project and the site selection should subsequently be tested using numerical modeling. Within this, a correct definition of the hydrogeologic setting should be considered in the model. Numerical modeling might assist in evaluating the potential impact of the project on groundwater levels surrounding the project, and might also be used to evaluate the potential fate and transport of contaminants.

Finally, only through pilot projects, the actual effect of the MAR project might be evaluated. This would open opportunities for the study of different techniques to enhance infiltration, and the use of different media to preclude aquifer contamination and treat the water used.

### 9.4. References


Chapter 9. General discussion


Chapter 9. General discussion


Martin, R., 2013. CLOGGING ISSUES ASSOCIATED WITH MANAGED AQUIFER RECHARGE METHODS 214.


Overton, I.C., Doody, T.M., 2008. Ecosystem changes on the River Murray floodplain over the last 100 years and predictions of climate change, in: Taniguchi, M., Burnett, W.C., Fukushima, Y., Haigh, M., Umezawa, Y. (Eds.), From headwaters to the ocean-hydrological changes and watershed management. Taylor and Frances Group, London.


