# Deriving a dataset from the New Zealand National Soils Database (NZ-NSD) for use in Soil and Water Assessment Tool (SWAT) Modelling

A. Parshotam
Aqualinc Research Ltd
aroon@aqualinc.co.nz

## Abstract

Pre-processed data from digital soil databases and maps provide necessary inputs to numerous biophysical, process-based enviro-agronomic models. Data from the New Zealand National Soils Database (NZ-NSD) was pre-processed to offer a New Zealand-level soils dataset in a format ready to be used for models, such as the Soil and Water Assessment Tool (SWAT) model. Variables required for SWAT include soil texture, bulk density, hydrological groups, available water capacity, saturated hydraulic conductivity, erodibility factors, carbon and soil albedo. NZ-NSD were joined to the New Zealand Fundamental Soils Layer (FSL) soil polygons by New Zealand Classification (NZSC) 'codes', within ArcGIS. Alternative codes from the FSL were assigned where there was no soil with corresponding codes in the NSD. These in turn were given a choice of soil sample IDs from the NSD, based on the data rating. Methods for quality control and for filling missing data are developed and documented, together with the issues and problems involved with the join process. The methodology may be expanded to include other soil variables of interest.

## Introduction

Many countries have publicly available pre-processed soils data from their respective national soils databases for use in physically-based hydrological, transport with crop-growth models, such as the Soil and Water Assessment Tool (SWAT) model. Deriving these data can be a laborious and time consuming task. Cordeiro *et al*. (2017), compiled a dataset from over 14 000 unique soils in the Soil Landscapes of Canada (SLC) database, with 85% of these soils having complete information in the database. A Pan European SWAT soil dataset was developed from the European Soil Database (ESDB) by Chambel-Leitão *et al*. (2012) for the EU *MyWater* Project. The Digital General Soil Map of the United States or STATSGO2 database is distributed in state/territory and national extents and designed for state level, and river basin applications. The STATSGO2 database is a revised STATSGO and is distributed by the Natural Resources Conservation Service (NRCS). It is freely available at http://soils.usda.gov/survey/geography/statsgo. The Soil Survey Geographic Database (SSURGO) is also distributed by NRCS and has the most detailed level of soil mapping and is structured on a county basis. It is designed for use by landowners, townships, and counties for natural resource planning and management. It consists of spatial and tabular data files. It is freely available at http://soildatamart.nrcs.usda.gov/. ArcMap tools were developed to automatically convert 1) STATSGO (1994)-State Soil Geographic Databases, 2) The U.S. General

Soil Map or STATSGO2 (2006), and 3) SSURGO2-Soil Survey Geographic Database into an ArcSWAT compatible form (Sheshukov *et al.,* 2009). The STATSGO and STATSGO2 databases are included in the SWAT database as a default.

In the absence of country databases, the FAO soils database has been used and datasets made available. For example, a SWAT Soil Database using FAO Soil and Terrain Database of East Africa (SOTER) data was prepared by Gies and Merwade is available at https://web.ics.purdue.edu/~vmerwade/education/fao_soil_tutorial.pdf. Soils information from World databases are available from the FAO soil portal at http://www.fao.org/soils-portal. This gives 1228 legacy maps from 80 countries around the world and, apart from Antarctica, New Zealand is not represented. The Harmonized World Soil Database is a 30 arc-second raster database with over 15 000 different soil mapping units that combines existing regional and national updates of soil information worldwide (SOTER, ESDB, Soil Map of China, World Inventory of Soil Emission Potentials (WISE)) with the information contained within the 1:5 000 000 scale FAO-UNESCO Soil Map of the World (FAO, 1971-1981). Gijsman *et al*. (2007) parameterized soil inputs for crop simulation models using the WISE database and 1125 soil profiles from around the world were converted into a format that could be used as input data to some commonly used biophysical computer models.

Soils databases are not in a consistent format across the participating countries. Data is often summarised from soil survey information and regularly updated for the scientific community to use. Legacy datasets (polygons and data) are often kept separate and not integrated into a 'supersoils' database. This means that different databases require different methodologies and approaches in creating SWAT soil datasets. The methodology is usually published and supplemented with detailed information on the derivation of data. Specialised software has been created in order to visualise Soils Data (e.g. *Soil Data Viewer* from http://docplayer.net/14707788-Soil-data-viewer-5-1-user-guide.html to use with US Soils databases).

The pre-1992 New Zealand National Soil Database (NSD), commonly called the NSD was developed by DSIR-Land Resources. It is available in the public domain but not widely publicised. Currently, it is the only dataset in the National Soils Data Repository (NSDR). The NSD contains much useful input data for process-based environmental models, such as SWAT, although data is rather sparse in some areas and there are some general issues with the database that can easily be patched. Parshotam *et al*. (1995) first utilised the NSD in a process-based modelling for a national assessment of soil carbon changes. Here, a soil map was produced by grouping together soil attributes using cluster analysis. Tate *et al*. (1993) used the concept of "similar" soils to group soils for New Zealand carbon reporting purposes.

The New Zealand Fundamental Soil Layer (FSL) originates from an 'expert derived' join of attributes measured in the NSD and polygons of the New Zealand Land Resource Inventory (NZLRI) using the dominant soil type/soil 'symbol' to join it to the NSD. These polygons in the FSL and the NZLRI and are in the most part, equivalent. The NZLRI maps an inventory of five physical factors (rock type, soil, slope, present type and severity of erosion, and vegetation) with a 'homogeneous unit area' recording the five physical factors simultaneously to a level of detail appropriate for presentation at a scale of 1:50 000. The NZLRI was not developed using soil profile data to delineate soil spatial polygons and uses some data from soil surveys. Although the FSL has been adopted as a useful national soil map, it is also considered to be a *pseudo-soil* map. The FSL contains spatial information for 16 key soil attributes. The attributes in the FSL were selected with consultation with stakeholders at the time and may not have considered the needs and requirements of process-based models, such as SWAT, and its precursors, which have been used in New Zealand for many years (see Parshotam and Robertson, 2018). The use of the FSL in its current form is therefore seen as a limitation. In this work, the NSD was joined to Fundamental Soils Layer (FSL) soil polygons by New Zealand Soil Classification (NZSC) codes. The numerous problems faced, solutions and workarounds are discussed.

This paper is about pre-processing the NSD database to offer a NZ-level soils dataset in a format ready to be used in process-based models such as SWAT. A summary description of the SWAT variables required is given. These variables include soil texture, bulk density, hydrological groups, available water capacity, saturated hydraulic conductivity, erodibility factors, carbon, and soil albedo for individual horizons in the soil profile. Further details of the approach are provided in Parshotam (2018).

## Methodology

The process of joining polygons from the FSL to soil profile data in the NSD for use in models such as the SWAT model involves two parts: 1) MS-Access database queries of the NSD to produce a list of soil identifiers with raw or derived data for individual horizons in the soil profile with soil texture, bulk density, hydrological groups, available water capacity, saturated hydraulic conductivity, erodibility factors, carbon, and soil albedo, and 2) linking FSL New Zealand Soil Classification (NZSC) polygons to relevant soil sample identifiers in the NSD.

The derived data may be done outside of MS-Access and imported into database. The MS-Access queries are not described in this work. The variables and source of data from relevant databases are summarised in Table 1. A brief description of the variables in the soil input file of the SWAT model is given below together with their derivation.

The NSD data were joined to the New Zealand Fundamental Soils Layer (FSL) soil polygons by New Zealand Classification (NZSC) ‘codes’, using ArcGIS. The code used was soil order/group/subgroup. New codes from the FSL were assigned where with a soil did not have a corresponding code in the NSD. These in turn were given a choice of soil sample IDs from the NSD, based on the quality of data. Methods for quality control and for filling missing data are developed and documented, together with the issues and problems involved with the join process.

A final excel spreadsheet is created where red highlighted entries indicate modelled data and not measurement data. Silt which is determined from sand and clay percent is not highlighted in red if sand and clay values are given. Error checking and fixing data in the NSD so that queries would perform better was documented.

A full description of the methodology is given by Parshotam (2018).

**Table 1:** Variables in the SWAT soil input file, and the source of data in tables from tables and fields in the NSD, and fields in the FSL.

| Variable name | Definition | Units | Geodatabase/ Database | Table | Relevant fields |
|---|---|---|---|---|---|
| SNAM | Soil name. The soil name printed in HRU summary tables [optional] | | FSL/NSD | | Either<br>1) Main soil/ Soil symbol except for Gisborne, or<br>2) NZSC for all of New Zealand<br>NSD SB sample identifiers |
| S5ID | [optional] | | NSD | | SB identifiers |
| HYDGRP | Soil hydrologic group (A, B, C, D) | | NSD | site | MIDS tables |
| SOL_ZMX | Max rooting depth [required] | mm | FSL | | PRD |
| | | | NSD | | Maximum depth of soil profile |
| ANION_EXCL | Fraction of porosity (void space) from which anions are excluded [optional] | -- | -- | -- | -- |
| SOL_CRK | Crack volume potential of soil [optional] | - | NSD | moisture | COLE,<br>Default=0.5 |
| TEXTURE | Soil texture [optional] | | NSD | Site<br>horizon | Type qualifier (texture)<br>Texture |
| SOL_Z(layer #) | Depth from soil surface to bottom of layer | mm | NSD | horizon<br>sample | Horizon top<br>Horizon Base<br>Sample top<br>Sample base |
| SOIL_BD (layer#) | Moist bulk density [required] | $g/cm^3$ | NSD | moisture | Dry bulk density, Fine earth BD |
| SOL_AWC (layer #) | Available water capacity of soil layer [required] | v/v% | NSD | moisture | Total available water |
| SOL_K (layer #) | Saturated hydraulic conductivity [required] | mm/hr | NSD | particle size<br>chemistry | regression equations from clay, sand, silt, CEC, OM<br>Pedotransfer functions |
| SOL_CBN (layer #) | Carbon content [required] | % soil mass | NSD | chemistry | Carbon |
| CLAY (layer #) | Clay content [required] | % soil mass | NSD | particle size | FE2 (FE=fine earth) |
| SILT (layer #) | Silt content [required] | % soil mass | NSD | particle size | Si |
| SAND (layer #) | Sand content [required] | % soil mass | NSD | particle size | Sa |
| ROCK (layer #) | Rock fragment content [required] | % total mass | NSD | particle size | 100-WS2 (WS2=whole soil less than 2mm) |
| SOL_ALB (top layer) | Moist soil albedo (top layer) [required] | --- | NSD | horizon | Matrix colour code (Munsell) |
| USLE_K (top layer) | USLE equation soil erodibility factor [required] | 0.013 (tonnes $m^2$ h)/$m^3$ tonne cm) | NSD | particle size | Silt, clay, sand, Org C |

## Soil variables of interest

The variables which follow are needed in the SWAT model.

**SNAM**

This is the soil name. There is the option of using soil ‘symbols’ (e.g. OkS for Okareka steepland soils), or New Zealand classification (NZSC) ‘codes’ for soil order/group.subgroup (e.g. BOT for Typic Orthic Brown Soil). NZSC codes were used in the following work for the following reasons: 1) the FSL is incomplete in the Gisborne area, and has to be treated in this area as an outlier, so one does not have the option of using dominant soil ‘symbols’ for a consistent system across New Zealand, 2) one does not require expert knowledge of a soil and the reference book by Hewitt *et al*. (2010) may be quite useful used to match similar soils if data is missing. The dominant (unique) NZSC name DOMNZSC in the FSL is chosen instead of the NZSC name, which includes combinations of soils (e.g. ROT+ROW).

**HYDGRP**

Soil hydrologic group (A, B, C or D). Note that there are no dual groups required in the SWAT model. A simplified Hydrologic Soil Group classification table (MPCA, 2013) which uses only texture information was used. The field texture was taken from the *horizons* table, in the NSD. Organic Soils were seen as an exception and classified group C for convenience, based on their usual location. Where there was no texture data filled in, twelve soil textures were defined using the NRCS texture calculator using clay, sand and silt percentages from the *nsd_particlesize* table in the NSD.

**SOL_ZMX**

Maximum rooting depth of the soil profile (mm). If no depth is specified, the model assumes the roots can develop throughout the entire depth of the soil profile. The PRD (plant rooting depth) should be taken from the FSL, if there were more clarity about its derivation. It does not appear to be in the NSD and perhaps is an artefact of migrating the NZLRI to FSL.

For the current purposes, the maximum depth of all horizons where data was available is taken to be the maximum rooting depth.

**ANION_EXCL**

The fraction of porosity (void space) from which anions are excluded. By default, this is taken to be 0.5 in the SWAT model.

**SOL_CK**

Potential or maximum crack volume of the soil profile expressed as a fraction of the total soil volume. A crack flow submodel was incorporated into SWAT and it is considered important in some soils and can contribute to movement of solutes to aquifers (Arnold et al., 2005). The crack volume may be derived from the Coefficient of Linear Extensibility (COLE) in the NSD, where available.

**TEXTURE**

There are three methods by which this may be obtained from the NSD.

1) NSD *site* table, *Type qualifier* field

2) NSD *horizon* table, *Texture* field

3) Defined according to clay, sand and silt% in the NSD using the NRCS texture calculator or Excel macros available online.

**SOL_Z (layer #)**

Depth from soil surface to bottom of layer (mm).

This is taken from the horizon table in the NSD. Every horizon in the NSD has a horizon top and horizon base, a sample top and a sample base. The sample top and base are a representative sample of the horizon top and base. Horizons are numbered 1, 2, 3, ... with associated lab letters A, B, C, … respectively. The sample top of the A horizon is taken as zero. In some instances, there is a negative value due to a litter layer, and the litter layer is not used, and the top layer is assigned zero. In some instances, there may be a gap between horizons due to only taking representative samples in a horizon (e.g. 0-7, 8-13, 13-20, 22-50, etc.). This is ignored and only the horizon base is taken. In some instances, a horizon is missing altogether in the NSD (i.e. no data, no horizon), in which case the last layer with data is taken as the bottom horizon.

**SOL_BD (layer #)**

The moist bulk density (g/cm$^3$) is taken to be the dry bulk density in the NSD, as their definitions are equivalent. Regression equations based on soil carbon and bulk density relationships was used where there was missing data.

**SOL_AWC (layer #)**

The plant available water, also referred to as the available water capacity of the soil (mm water/mm soil). In the NSD, available water is calculated as the difference between water content at tensions 15 bar and 0.33 bar, with units (v/v%).

**SOL_K (layer #)**

Saturated hydraulic conductivity (mm/h). This was derived using a spreadsheet form of the SPAW-Soil Characteristics model given in Saxton and Rawls (2006). It is used to simulate soil water tension, conductivity and water holding capability based on the soil texture, with adjustments to account for gravel content, compaction, salinity, and organic matter.

**SOL_CBN (layer #)**

Organic carbon content (% soil mass), taken directly from the NSD.

**SOL_CLAY (layer #)**

This is the clay fragment content (% of total mass), taken directly from the NSD.

**SOL_SAND (layer #)**

This is the sand content (% of total mass), taken directly from the NSD.

**SOL_SILT (layer #)**

The silt content (% of total mass), taken directly from the NSD. Where it is missing it is calculated from clay (%), sand (%) and rock (%)

**SOL_ROCK (layer #)**

The rock fragment content (% of total mass), taken directly from the NSD.

**SOL_ALB (top layer)**

The *moist* soil albedo is derived from Munsell® colour codes in the NSD.

**USLE_K (top layer)**

The Universal Soil Loss Equation (USLE) soil erodibility factor, is a quantitative description of the inherent erodibility of a soil. It is derived from organic matter (%), silt (%), sand (%), clay (%).

## NSD-FSL join

The following has to be considered, in the first instance:

1. The use of as many NSD site identifiers (and corresponding good 'quality' soils data, with very little missing data) as possible, and a high degree of one-to-one correspondence between soil classification 'name' and site identifiers.
2. The NSD has a lot of missing data; some tables in the NSD were derived from multiple sources are not linked, and there are numerous spelling mistakes which confound a successful database query involving text.
3. The FSL is incomplete in the Gisborne area, and has to be treated in this area as an outlier, so one does not have the option of using dominant soil 'symbols' for a consistent methodology across New Zealand.

A summary of unique classes and counts in those classes in the FSL and NSD that give viable options for a join is given by Parshotam (2018). Names (IDs) based on local soil names from soil surveys are common and convenient to explain to land owners and clients, and are most commonly used internationally. In New Zealand, have studies found local soil 'symbols' convenient to describe the effect of sediment generated from soils (Parshotam 2008; Parshotam *et al.*, 2009).

In the current work, a join was made possible by linking the NSD using NZSC codes. Since there were so many factors (e.g. several NSD site identifiers which corresponded to a single polygon in the FSL) to consider, a set of rules had to be developed and the process manually implemented, i.e. not automated, and by judgement. The publication New *Zealand Soil Classification* by Hewitt (2010) was used as a reference book to match soils so that no 'expert' knowledge of local soils became necessary. However, expert knowledge of local soils was necessary in joining the NSD to the NZLRI when matching soil symbols to derive the FSL, largely as a result of having to choose relevant soil symbols within a correlation set. The following set of rules were created:

1. The NSD is a database of point data and associated with every point in the NSD is a NZ Revised subgroup class, which is representative of a spatial area with the same soil.
2. In choosing site numbers, sites were chosen to be representative with as much data as possible with no further processing necessary. Sites that did not need missing data to be estimated and added were given priority and a high quality value. A priority was to choose sites with missing data that could easily be filled. If a site did not have coordinates, it was ignored. If there were two sites to choose and one had a litter layer which started with -20cm, the other which did not have a litter layer was chosen. Spelling mistakes that confound database querying were just considered an artefact of the database and not "fixed". For example, there is a "RF?" DOMNZSC class in the FSL which was not assigned RFA because there was no justification in doing so and was assigned a new class, RFU, for 'unknown'.

3. When assigning new codes for classes that exist in the FSL but not in the NSD, the following rule was followed. The original code was defined from the New Zealand Soil Classification (Hewitt, 2010). Another soil from the same class which appeared to be similar was chosen, and if there were several to choose from, the soil with the greatest amount of data was chosen. When it was difficult, such as the case with the Anthropic Soils AFST and ATT, a soil that was correlated was chosen instead from Hewitt (2010).
4. Sites with many horizons and a lot of data were given priority.
5. Sites were selected irrespective of whether they had a Munsell® colour code or not.
6. Any data could be missing. There were some samples with missing horizons (e.g. SB09303 missing F) and SB09506 starts at horizon C with no A or B).
7. A star scoring system was given to rate the data from one star to five stars. Five stars was assigned to a site with complete data and fewer stars were assigned to sites with missing data.A comment added to record which data was missing.
8. Later numbered samples (e.g. from the SB10000 series rather than the SB08000 series) often had more data and so the data was sorted by date when identifying.

Table 2 shows the results of applying this reclassification system. The highlighted blue cells in the FSL.DOMNZSC represent soils that are not in the NSD and need to be assigned a new Subgroup according to Hewitt (2010).

Table 3 shows the primary choice and a secondary choice is given along with a rating of the quantity and quality of data, and a justification comment.

**Table 2:** Part of a table showing the NSD NZ Revised subgroup symbols and corresponding full NSD Revised Soil classes. The highlighted blue cells in the FSL.DOMNZSC represent soils that are not in the NSD which need to be assigned new codes according to Hewitt (2010). The full table is given in Parshotam (2018).

| NSD Site.NZ Revised subgroup | NSD Site.NZ Revised Soil class | FSL.DOMNZSC | New code to replace FSL.DOMNZSC that do not exist in NSD | comment/justification for replacing; with page number reference in Hewitt (2010) |
|---|---|---|---|---|
| | | AFST | WGFU | Fill Anthropic. p.50; correlated with entisols, gley-raw soils |
| | | ATT | WGFU | Truncated Anthropic Other soils. P.49; fill anthropic. p.50; correlated with entisols, gley-raw soils |
| BAM | Mottled Acid Brown Soil | BAM | | |
| BAMP | Mottled-placic Acid Brown Soil | | | |

**Table 3:** Part of a table showing the primary choice and a secondary choice is given along with a rating of the quantity and quality of data and a justification comment. The "FSL.DOMNZSC" column represents the dominant soil in the Fundamental Soils Layer (FSL) classified by the New Zealand Soil Classification (NZSC). The blue highlighted cells represent soils which do not have a corresponding NZ Revised soil class in the NSD. The "FSL.DOMNZSCamended" column represents similar soils from Hewitt (2010), with a good match of data from the NSD. A primary choice and a secondary choice are given along with a rating of the quantity and quality of data. The full table is given in Parshotam (2018).

| FSL.DOMNZSC | FSL.DOMNZSC amended | Primary choice | Secondary choice | Data rating | comment |
|---|---|---|---|---|---|
| AFST | AFST | SB09307 | SB09308 | *** | no moisture, no PS |
| ATT | ATT | SB09307 | SB09308 | *** | no moisture, no PS |
| BAM | BAM | SB09870 | | **** | no moisture |
| BAO | BAO | SB08979 | | ** | no moisture, no PS |
| BAP | BAP | SB09330 | | **** | no moisture |
| BAT | BAT | SB10046 | SB09917 | ***** | |
| BFA | BFA | SB09447 | SB09448 | ***** | |
| BFAL | BFA | SB09447 | | ***** | |
| BFC | BFA | SB09947 | | ***** | |
| BFL | BFL | SB10138 | | ***** | |

## Results

Table 4 gives a screenshot of an Excel spreadsheet showing part of the full dataset. Note that the red indicates 'modelled' data and not 'measurement' data in the NSD, which is in black. This data represents data from only the top horizon (NLAYERS=1) but data from other horizons are available and may be included.

**Table 4:** A screenshot of the Excel spreadsheet of part of the full dataset. Note that the red indicates 'modelled' data and not 'measurement' data in the NSD, which is in black.

| SNAM | S5ID | CMPPCT | NLAYERS | HYDGRP | SOL_ZMX | ANION_EXCL | SOL_CRK | TEXTURE | SOL_Z1 | SOL_BD1 | SOL_AWC1 | SOL_K1 | SOL_CBN1 | CLAY1 | SILT1 | SAND1 | ROCK1 | SOL_ALB1 | USLE_K1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AFST | SB09307 | 100 | 1 | D | 100 | 0.5 | 0.5 | silty clay | 100 | 1 | 0.17 | 26.94 | 9.4 | 34 | 54 | 12 | 0 | 0.162 | 0.229772 |
| ATT | SB09307 | 100 | 1 | D | 100 | 0.5 | 0.5 | silty clay | 100 | 1 | 0.17 | 26.94 | 9.4 | 34 | 54 | 12 | 0 | 0.162 | 0.229772 |
| BAM | SB09870 | 100 | 1 | D | 100 | 0.5 | 0.5 | silty clay loam | 100 | 1.06 | 0.14 | 17.87 | 8.2 | 42.82 | 48.04 | 9.14 | 0 | 0.162 | 0.232552 |
| BAO | SB08979 | 100 | 1 | B | 50 | 0.5 | 0.5 | silt loam | 50 | 1.05 | 0.22 | 48.52 | 6 | 14 | 65 | 21 | 0 | 0.093 | 0.225805 |
| BAP | SB09330 | 100 | 1 | B | 190 | 0.5 | 0.5 | silt loam | 190 | 0.86 | 0.23 | 103.06 | 7.9 | 8 | 62.3 | 29.7 | 0 | 0.093 | 0.203983 |
| BAT | SB10046 | 100 | 1 | B | 180 | 0.5 | 0.5 | silt loam | 180 | 0.395 | 0.15 | 41.15 | 11.6 | 24 | 54.65 | 21.35 | 7 | 0.093 | 0.196693 |
| BFA | SB09447 | 100 | 1 | B | 80 | 0.5 | 0.5 | silt loam | 80 | 1.01 | 0.318 | 14.36 | 5.3 | 34 | 53.65 | 12.35 | 0 | 0.162 | 0.226471 |

## Discussion

Deriving a dataset from national databases for use in process-based models such as SWAT can be a laborious and time-consuming task. However, datasets of this form are useful to a variety of research communities including hydrological, agricultural, agronomic, and water quality modellers. These are often made publicly available at (e.g. SWAT usersoil Canada https://www.earth-syst-sci-data-discuss.net/essd-2017-66/). When comparing the New Zealand and Canadian data there is 3.5 times as much data in NSD as in the SLC on a $km^2$ basis.

The original process of joining the NSD and the NZLRI to form the FSL was not documented. Part of this work was to understand how the FSL was created from the NZLRI in order to add further attributes easily if necessary. A rough analysis (by observing repeated values of attributes) shows that much less than a third of the data in the NSD might actually be used in the join to create the FSL. In the current work, a lot of data (85%) was not used to derive the final dataset, and this includes good quality data. Data were used from only about 200 out of 1500 samples in the NSD. A number of sites had a lot of excellent data to represent this class (e.g. the soil order/group/subgroup codes given by PUJ, GOI, GAY, BAMP) but they are not utilised, because the process only takes one representative sample from the NSD. There is still a great need for more accurate soil profile data to build up a good working database. Much pre-1992 legacy data on 'coloured' report cards exists that needs to be added to the database, and made publically available.

The use of S-map, the digital soil map which supersedes the FSL, was explored and S-map fact sheets were found to be useful for cross-checking hydrologic groups for Organic Soils. S-map makes use of some soil legacy maps (from the 1950s) to delineate polygons in some areas around the country using a 'non-local' classification system. Much of the soil attributes data required by SWAT is not currently available in S-map, which better supports nutrient budget models than for informing models like SWAT. Although S-map online individual fact sheets or soil reports were of some use, query-able databases would be of far greater use. S-map only covers a third of the country (ie. incomplete national coverage) and a consistent methodology was required for the whole country. S-map polygons need to be purchased – they are not freely available. A lot of data was given as a range (e.g. 0-40% stones, clay 0-20%) and represented an area rather than a representative sample. And there was no system for determining what was modelled or 'filled' data from pedo-transfer functions and/or measurement data. In view of all these factors, the FSL, although not ideal was available in the public domain and favoured.

New methods that make use of all data in the NSD to create soil data in the form of raster data should be explored. For example, there are methods to interpolate point data and represent these spatially, e.g. The Soil-Landscape estimation and Evaluation Program (SLEEP) by FAO and SWAT modelling scientists was developed to predict spatial distribution of soil attributes for environmental modelling (Ziadat *et al.*, 2015). The SLEEP tool spatially interpolates measured soil attributes and produces a continuous representation of soil. It uses location and measured soil attribute data, DEM (Digital Elevation Model), Red band of the satellite image and Infra-red band of the satellite image to produce a soil attribute prediction (in Table and Raster Format) and a SLEEP-SWAT Format conversion. The SLEEP tool divides a watershed or area into different zones or "facets" based on the average slope parameters and then derives a model for each facet relating the soil attributes to different terrain and environmental attributes. Where more detailed soil survey maps are not available, data on geology, topography, vegetation, and climate are assembled and related to satellite images. SLEEP uses measured soil properties (e.g. soil depth and percentage content of clay, silt, sand, stone and organic matter) at different locations in a catchment along with the geographical co-ordinates of the measurement locations, to produce the spatially distributed soil properties for the whole catchment in the form of raster data, and for different soil horizons. These catchment distributed soil properties may be converted to database form. The advantage with this approach is that it does not rely on specialised knowledge of soil classification systems, which can be a hindrance.

## Conclusions

Pre-processing critical soils data can be laborious and time-consuming but a necessary process in order to prepare inputs required in numerous process-based enviro-agronomic models.

Data from the New Zealand National Soils Database (NSD) was pre-processed to offer a New Zealand-level soil dataset in a format ready to be used for models, such as the Soil and Water Assessment Tool (SWAT) model. Variables required for SWAT include soil texture, bulk density, hydrological groups, available water capacity, saturated hydraulic conductivity, erodibility factors, carbon and soil albedo.

NZ-NSD were joined to the New Zealand Fundamental Soils Layer (FSL) soil polygons by New Zealand Classification (NZSC) 'codes' (soil order/group/subgroup), within ArcGIS. Alternative codes from the FSL were assigned where there was no soil with corresponding codes in the NSD. These in turn were given a choice of soil sample IDs from the NSD, based on the data rating. Methods for quality control and for filling missing data are developed and documented, together with the issues and problems involved with the join process. The methodology may be expanded to include other soil variables of interest. The method does not make use of a lot of good quality data in the NSD and alternative methods that make full use of all data in the NSD to create soil data in the form of raster data should be explored.

## Acknowledgements

This soil dataset was necessary for preparing input data to the land, freshwater and marine erosion and sediment model, funded by the Ministry for the Environment.

## References

Arnold, J.G., Potter, K.N., King, K.W. and Allen, P.M. 2005. Estimation of soil cracking and the effect on surface runoff in a Texas Blackland Prairie watershed. *Hydrological Processes.* Vol 19: 589-603. https://doi.org/10.1002/hyp.5609

Chambel-Leitão, P., Almeida, C., Weynants, M., Tóth, G., Tóth, B. and Neves, R. 2012. A European SWAT soil database. Paper presented at the 2012 International SWAT Conference and Workshops, New Delhi, India.

Cordeiro, M.R.C., Lelyk, G., Kröbel, R., Legesse, G., Faramarzi, M., Masud, M.B. and McAllister, T. 2017. Deriving Canada-wide soils dataset for use in Soil and Water Assessment Tool (SWAT). PANGAEA, https://doi.org/10.1594/PANGAEA.877298, Supplement to: Cordeiro, MRC et al. (2017): Deriving a country-wide soils dataset from the Soil Landscapes of Canada (SLC) database for use in Soil and Water Assessment Tool (SWAT) Simulations. Earth System Science Data Discussions, 30 pp, https://doi.org/10.5194/essd-2017-66

Gies, L. and Merwade, V. Creating SWAT Soil Database using FAO Soil and Terrain Database of East Africa (SOTER) Data. Available from https://web.ics.purdue.edu/~vmerwade/education/fao_soil_tutorial.pdf

Gijsman, A., Thornton, P.K. and Hoogenboom, G. 2007. Using the WISE database to parameterize soil inputs for Crop Simulation Models. *Computers and Electronics in Agriculture* 56(2): 85-100. https://doi.org/10.1016/j.compag.2007.01.001

Hewitt A.E. 2010. New Zealand Soil Classification. 3rd edn. Landcare Research Science Series No. 1. Lincoln, Manaaki Whenua Press.

Minnesota Pollution Control Agency, (MPCA). 2013. Minnesota Stormwater Manual, Overview of Basic Stormwater Concepts. Available online at: http://stormwater.pca.state.mn.us/index.php/Main_Page

Parshotam, A., Tate, K. and Giltrap, J. 1995. Potential effects of climate and land use change on soil carbon and CO2 emissions from New Zealand's indigenous forests and unimproved grasslands. Weather and Climate. 152, 47-45. Weather and Climate, http://www.metsoc.org.nz/sites/default/files/journals/15/2/articles/1995_152_47-56_parshotam.pdf

Parshotam A. 2008. Central Waitemata Harbour Contaminant Study. GLEAMS model results for rural and earthworks sediment loads from the catchment. Prepared by DHI Water & Environment Ltd for Auckland Regional Council. Auckland Regional Council Technical Report 2008/041.

Parshotam, A., Wadhwa, S. and Mullan, B. 2009. Tauranga Harbour Sediment Study. Sediment Load Model Implementation and validation. *NIWA report number HAM2009-007*. Prepared for Environment Bay of Plenty.

Parshotam, A. 2018. New Zealand SWAT: Deriving a New Zealand-wide soils dataset from the NZ-NSD for use in the Soil and Water Assessment Tool (SWAT) model. Report prepared for Ministry for the Environment. Project: WL18033. Aqualinc Research Limited.

Parshotam, A. and Robertson, D. 2018. Chapter 2. Modelling for Catchment Management. In: Lake Restoration Handbook. A New Zealand Perspective. Springer. In press.

Saxton, K.E. and Rawls, W.J. 2006. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal* 79: 1569-1578.

Sheshukov, A., Daggupati, P., Lee, M.C. and Douglas-Mankin, K.R. 2009. ArcMap Tool for Pre-processing SSURGO Soil Database for ArcSWAT. Presented at the 5th SWAT International Conference in Boulder, CO – 5-7 August 2009, https://swat.tamu.edu/docs/swat/conferences/2009/C2-Sheshukov.pdf

Tate, K.R., Giltrap, D.J., Claydon, J.J., Newsome, P.F., Atkinson, I., Taylor, M.D. and Lee, R. 1997. Organic carbon stocks in New Zealand's terrestrial ecosystems. *Journal of the Royal Society of New Zealand* 27:3, 315-335, DOI: 10.1080/03014223.1997.9517541

Ziadat, F., Dhanesh, Y., Shoemate, D., Srinivasan, R., Narasimhan, B. and Tech, J. 2015. Soil-Landscape Estimation and Evaluation Program (SLEEP) to predict spatial distribution of soil attributes for environmental modelling. *International Journal of Agricultural and Biological Engineering* 8(3): 1-15, https://swat.tamu.edu/media/115182/sleep_ijabe.pdf