Cheryl Porter
University of Florida
2020-05-04

How to use an AgMIP formatted ACEB or JSON file

AgMIP data formats were developed to store and transfer data related to field crop experiments, farm surveys, and other agronomic data in an efficient, flexible format (Porter et al, 2014). AgMIP crop modeling protocols promote the use of multiple models for research applications (Rosenzweig et al, 2012). The AgMIP data products were originally developed to allow data from diverse sources to be harmonized using a common vocabulary, to be combined with expert knowledge, and to be translated equivalently to the model-specific formats used by individual crop models (Figure 1).

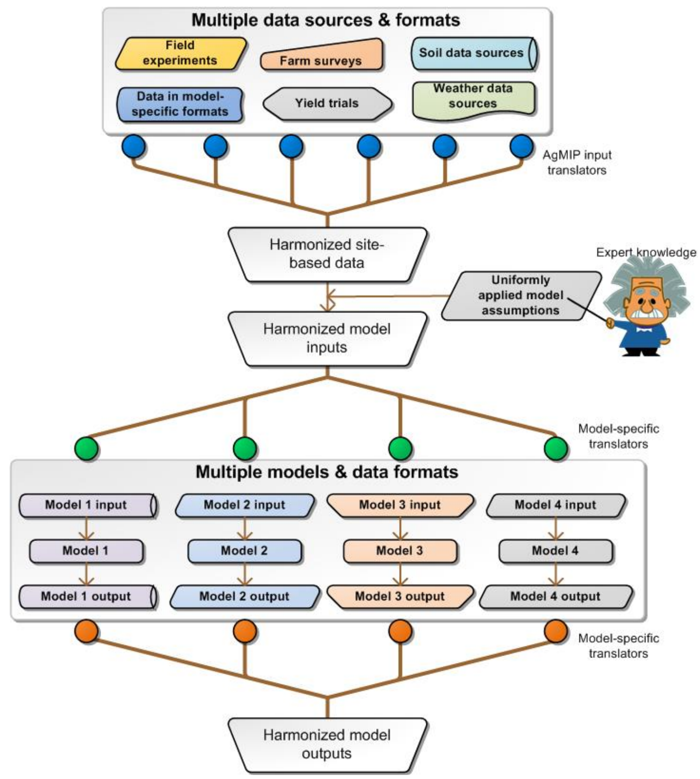

Figure 1. AgMIP data interoperability tools were developed to harmonize data from diverse sources with a standard vocabulary, to allow data to be combined with expert knowledge, and to translate the combined data stream to model-specific formats for multi-model activities.

Data format

The AgMIP harmonized data format is referred as ACE, or AgMIP Crop Experiment format. These semi-structured AgMIP data are stored in key-value pairs, where the key corresponds to an ICASA variable (www.tinyurl.com/icasa-mvl) which inherits a definition and a unit. AgMIP data may be available in either “JSON” or “ACEB” formats. The “JSON” (JavaScript Object Notation) suffix is a recognized data formatting standard. Many JSON readers are publicly available and allow easy parsing and processing. The “ACEB” format, or ACE binary format, is simply a GZipped (compressed) JSON file. AgMIP has developed tools for viewing ACEB data and for translating AgMIP formatted data to model-specific formats for several crop models.

As shown in Figure 2, AgMIP data files consist of three parts, experiments, weathers, and soils. Relational linkages occur through use of “soil_id” and “wst_id” variables. Data sub-structures are used for management event data. For example, in Figure 2, a fertilizer event has been expanded to show date of application, the nitrogen amount applied, and other details. Data sub-structures are also used for soil layer data and for time series data, such as weather or field observations of plant growth characteristics.

AgMIP Data Tools

AgMIP data interoperability tools can be freely downloaded from the AgMIP toolshed (http://tools.agmip.org/). The more useful tools for most Ag Data Commons users are:

ACEB Viewer (http://tools.agmip.org/acebviewer.php) – This tool allows ACEB files to be explored visually. Figure 3 shows an example of an ACEB file viewed with the ACEB Viewer.

QuadUI (http://tools.agmip.org/quadui.php) – This java software application can be run as a desktop application or from a command line. The application reads various data formats, including AgMIP JSON and ACEB files, and converts data to model-specific formats.

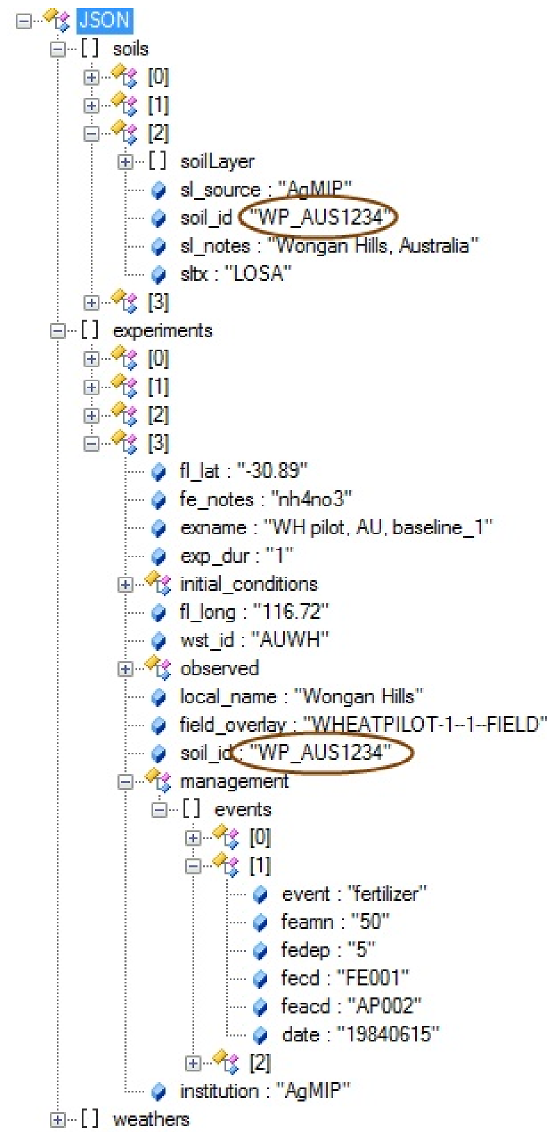

Figure 2. Snippet of an AgMIP JSON data

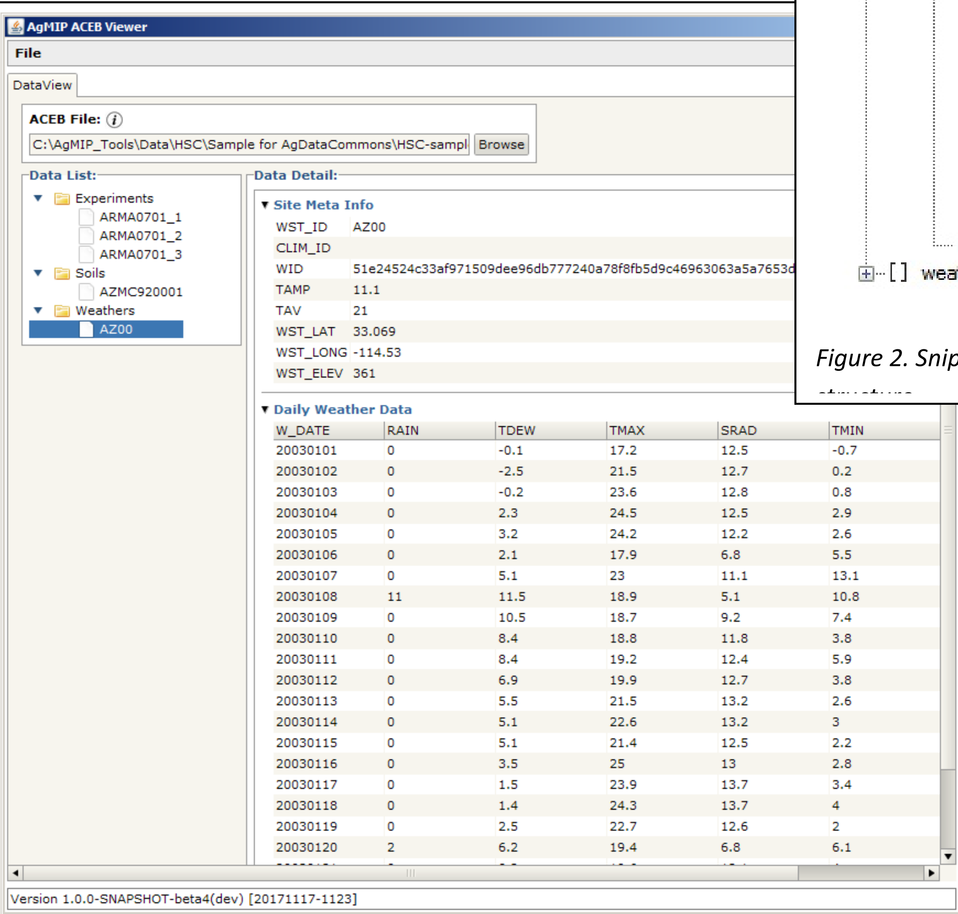

Figure 3. ACEB viewer showing part of Hot Serial Cereal weather data.

Example of translation for DSSAT Cropping System Model using QuadUI

Two sample data files are referenced in this document to illustrate how the AgMIP tools can be used to translate an ACEB file to DSSAT model format (as an example for model inputs).

NOTE: Other modeling formats can be selected through the QuadUI interface, although (*user beware!*) some of these translators have not been maintained and may only work for older versions of that model. Please contact the model developers and it@agmip.org if you need an updated translator.

The sample data used in this example are a subset of data for a field crop experiment conducted in Maricopa, Arizona, USA called the Hot Serial Cereal experiment (Kimball et al, 2016). The sample data files referenced herein are:

HSC-sample_4.5.aceb – AgMIP format file, ACEB format with experimental data collected in the field, including metadata, crop management, observations of crop growth and development, weather, and soils data.

Field_overlay_HSC_v4.5a.csv – DOME file to supply additional information required for modeling.

Figure 4 shows a screenshot of the QuadUI desktop application after a user has selected the *HSC-sample_4.5.aceb* file. (Note that QuadUI can also be used to translate raw data from Excel and other formats, where providing soils and weather data separately may be convenient, but in this case, the weather and soils data are included in the ACEB file downloaded from Ag Data Commons.)

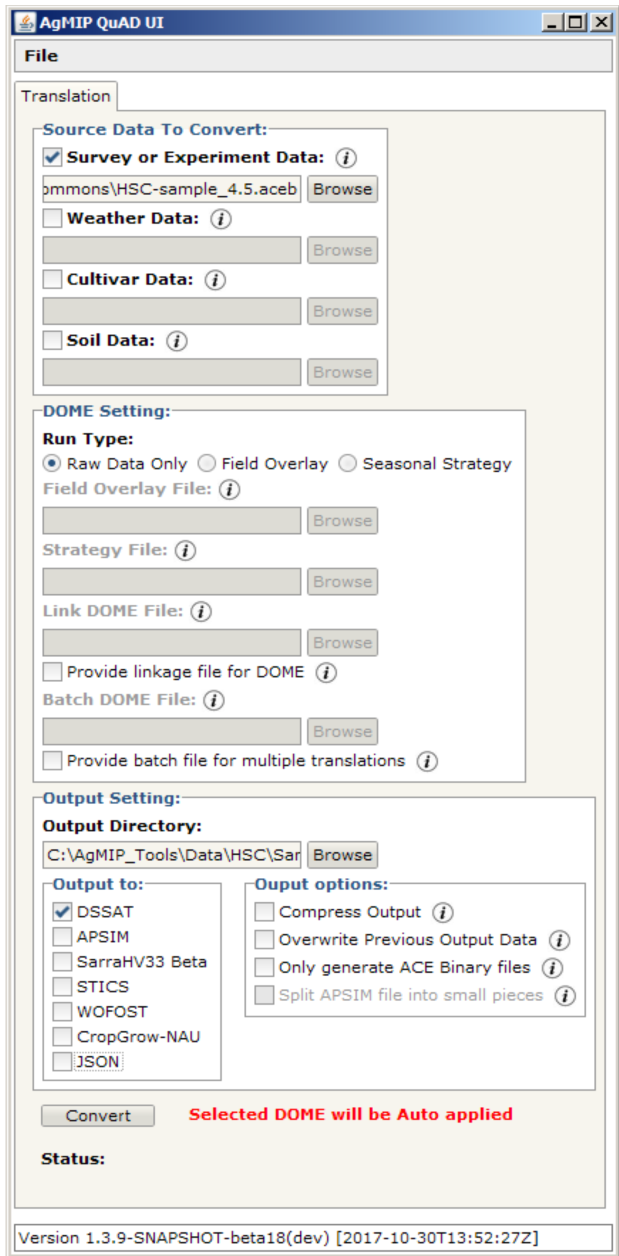

Figure 4. QuadUI configured with data file only

The user has selected to translate the data to DSSAT format. When the “Convert” option is selected, the data are translated to DSSAT format. Multiple crop modeling options may be selected simultaneously, and the model input files are created in separate folders for each model. In this case, a single folder for DSSAT will be created for the translated files.

When a user tries to run a simulation with the DSSAT files created, the model will generate errors because a few required model parameters are not present in the data collected in the field experiment. At this point, we need to introduce a separate data source that can supply these additional model parameters, determined or estimated by the modeler.

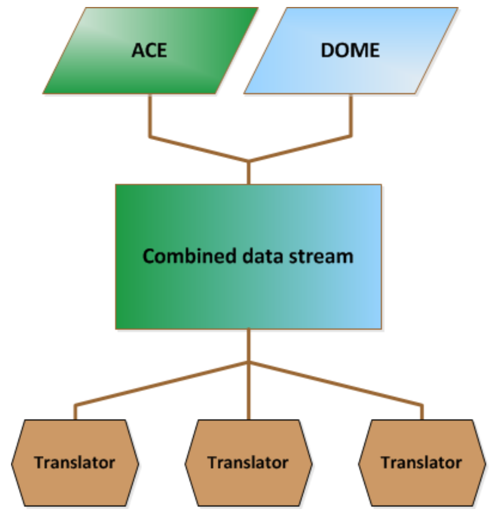

Figure 5. DOME data can supplement data recorded in the field to produce a complete set of data from crop modeling.

DOME files - When the end use of a dataset involves crop modeling, we often find that even detailed descriptions of experiments lack data required by a model. For example, perhaps the type of fertilizer or the row spacing was not recorded. A DOME (Data Overlay for Multi-model Export) file can supply additional information based on the best available knowledge of the cropping system being modeled. Figure 5 illustrates the concept. While both experimental data (ACE) and modeler assumptions (DOME) are combined to generate model inputs for the simulation, the original field-measured data are not modified.

DOME functions are available to allow calculation of modeling inputs such as initial water content, initial N distribution, and other functions (see http://research.agmip.org/display/itwiki/The+DOME).

DOMEs may also be used to generate scenarios for modeling future climates, adaptation strategies, sensitivity analyses, or other hypothetical scenarios. Addition of DOME data does not alter the original dataset, and multiple DOMEs can be combined to create complex scenarios.

Figure 6 shows a screenshot of QuadUI after adding the Field overlay DOME file. With the complete set of data required for modeling now provided, the simulations can be executed without error.

See the section of this document entitled: “How to Use a DOME” for more information on the structure and content of DOME data.

References

Kimball, B.A., J.W. White, G.W. Wall, M.J. Ottman 2016. Wheat Responses to a Wide Range of Temperatures: The Hot Serial Cereal Experiment. In: J. L. Hatfield, D. Fleisher, editors, Improving Modeling Tools to Assess Climate Change Effects on Crop Response, Adv. Agric. Syst. Model. 7. ASA, CSSA, and SSSA, Madison, WI. p. 33-44. doi:10.2134/advagricsystmodel7.2014.0014

Porter, C.H., C. Villalobos, D. Holzworth, R. Nelson, J.W. White, I.N. Athanasiadis, S. Janssen, D. Ripoche, J. Cufi, D. Raes, M. Zhang, R. Knapen, R. Sahajpal, K.J. Boote, J.W. Jones. 2014. Harmonization and translation of crop modeling data to ensure interoperability. Environmental Modelling and Software. 62:495-508. doi:10.1016/j.envsoft.2014.09.004.

Rosenzweig, C., J.W. Jones, J.L. Hatfield, A.C. Ruane, K.J. Boote, P. Thorburn, J.M. Antle,G.C. Nelson, C. Porter, S. Janssen, S. Asseng, B. Basso, F. Ewert, D. Wallach, G. Baigorria, and J.M. Winter. 2012. The Agricultural Model

Intercomparison and Improvement Project (AgMIP): Protocols and Pilot Studies. Ag. For. Meteor. 170:166-182 http://dx.doi.org/10.1016/j.agrformet.2012.09.011.

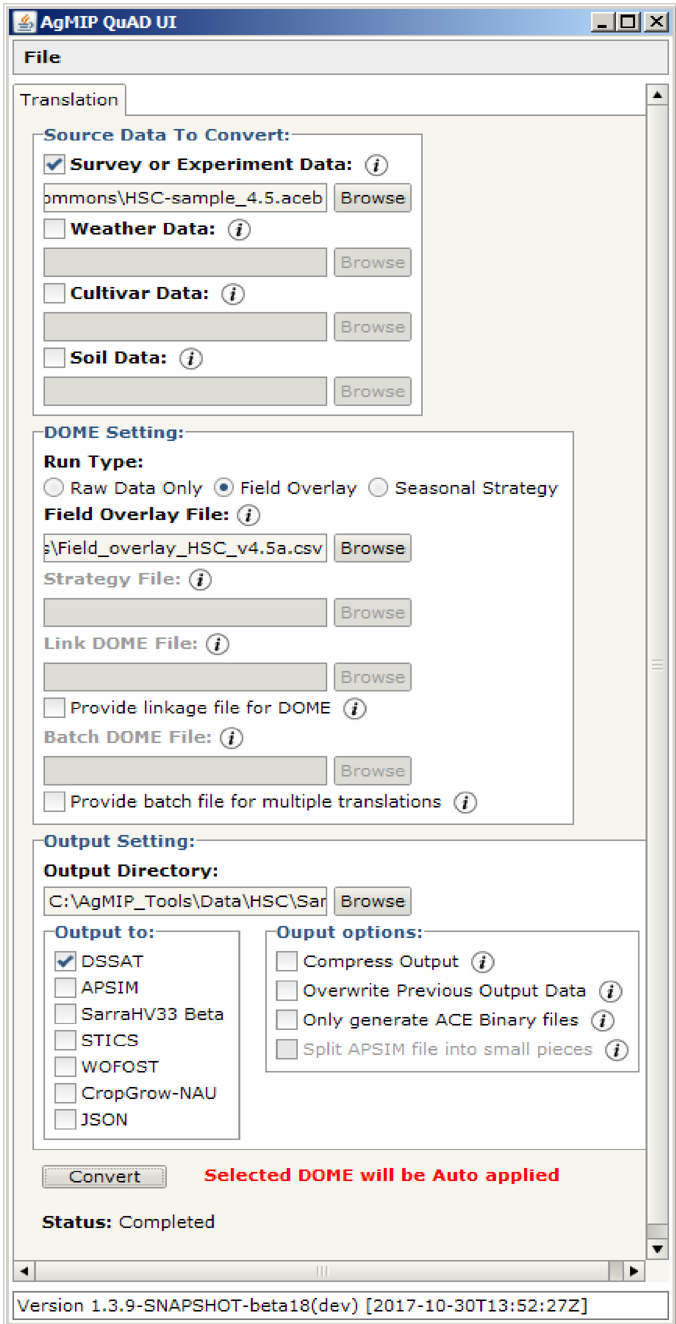

Figure 6. QuadUI with experimental and DOME data

How to create a DOME file

DOME files - A DOME (Data Overlay for Multi-model Export) file supplies additional information to supplement data collected in a field experiment. Addition of DOME data does not alter the original dataset and multiple DOMEs can be combined to create complex scenarios for simulation.

DOME functions are available (see http://research.agmip.org/display/itwiki/The+DOME) to estimate modeling inputs such as initial water content, initial N distribution, and planting date.

There are three kinds of DOME files:

- Field overlay DOMEs add information to supplement a baseline simulation. These data are used in addition to data collected in the field and are provided solely to complete model input data required for this baseline simulation.
- Seasonal Strategy DOMEs allow a single season of data to be used with multiple years of weather data to analyze scenarios based on seasonal weather variability. These DOMEs are used in conjunction with Field overlay DOMEs. Additional hypothetical management information may be introduced to mimic adaptations or other scenarios.
- Batch DOMEs allow loops of simulations to be done for sensitivity analyses, climate change studies, and other simulation series.

The various parts of a DOME file are explained below. This is not an exhaustive explanation of all capabilities of DOMEs, but is meant to give the user a feel for the way DOME files are constructed and some of the capabilities.

As an example, we will look at the Field overlay DOME used for the Hot Serial Cereal experiment, described in the previous section. The screenshots are equivalent to the data in the *Field_overlay_HSC_v4.5a.csv* file, but have been put into an Excel spreadsheet for a simpler view of data definitions.

General notes on the DOME file structure:

- Comments are denoted with "!" and are not processed by the translators. Fields beginning with "!" are ignored. Lines with "!" in column 1 are entirely ignored.
- Lines of data with "&" in column 1 are processed by the translators.
- The second column of every data line contains operators: INFO (for metadata), FILL (to fill in missing values), or REPLACE (to replace all values with the data provided).
- The third column contains the name of the variable being provided.
- The fourth column contains either a value or a function name.
- All function names include "()" after the function name. Columns 5 through N after a function name contain function parameters.
- Functions are applied in order, so if variable B depends on variable A, then variable A must be specified prior to variable B.

Metadata section

The metadata section at the top of every DOME file contains the full description of the scenario being modeled. This section is used by AgMIP Regional Integrated Assessment Teams to carry consistent metadata throughout workflows to provide provenance to final model outputs. Fields that are not relevant to a simulation may be left blank, but as a minimum, the DOME region identification (REG_ID) must be filled.

!	Metadata for Overlay			! Comment
!	Comment	**Dome_Name**	**HSC-----FIELD**	! Auto-generated from following 6 fields.
&	INFO	REG_ID	HSC	! Region ID
&	INFO	Stratum		! Socioeconomic or geographic stratum ID
&	INFO	RAP_ID		! Representative Agricultural Pathway (RAP) ID
&	INFO	MAN_ID		! Management or scenario ID
&	INFO	RAP_VER		! RAP version
&	INFO	DESCRIPTION	FIELD	! Description of scenario

The DOME name is constructed from a concatenation of each metadata field (orange block above), separated by a dash. It is important to not use dashes in the field identifier text for this reason. The DOME name is used to link multiple DOMEs to the corresponding data elements.

Initial conditions

The initial conditions block is often needed in simulations if the soil water and nitrogen contents were not measured at the beginning of the simulation. Modelers must make their best evaluation of field conditions. Several DOME functions facilitate this process.

!	Initial conditions			
!	**Dome operator**	**Variable to be modified**	**Value or Function**	**Function arguments**
!		**Initial soil water content (mm3/mm3)**	**Function Percent available water**	**ICSW% - Percent of available water (%)**
&	FILL	ICH2O	PCTAWC()	50
!		**Initial soil N conc (ppm)**	**ppm**	
&	FILL	ICNO3	5	
&	FILL	ICNH4	2	

In this example, initial soil water content (ICH2O) was estimated using the PCTAWC function (percent available water). The user estimated that at the beginning of simulation that soil water content in each layer of the soil profile was 50% between lower limit and drained upper limit. See http://research.agmip.org/display/itwiki/PCTAWC+-+Percent+available+water for more information on this function.

The initial soil nitrate (ICNO3) and ammonium (ICNH4) values are provided as values. These variables are arrays and so the same value is used for every element of the array. In this case, initial nitrate concentration is set to 5 ppm and ammonium to 2 ppm in every soil layer.

The csv form of this initial conditions section would look like this:

```
&,FILL,ICH2O,PCTAWC(),50
&,FILL,ICNO3,5
&,FILL,ICNH4,2
```

Planting data. Planting and soil data are similarly filled with modeler-supplied parameters. Some of the values might be specific to a particular model. For example, in the Planting Event section, cultivar names specific to each model are provided. The model ID is used as a prefix for each. In this case, the "REPLACE" operator is used to provide cultivar names. These names will override any cultivar IDs in the data, if they are provided.

!	**Planting Event**		
!	**Dome operator**	**Variable to be modified**	**Value or Function**
!		**Planting material**	**code (S=dry seed)**
&	FILL	plma	S
!		**Planting distribution**	**code (R=rows)**
&	FILL	plds	R
!		**cultivar ID for models**	
&	REPLACE	dssat_cul_id	CI0001
&	REPLACE	apsim_cul_id	bolac
&	REPLACE	aquacrop_cul_id	wheat_v1
&	REPLACE	stics_cul_id	1
&	REPLACE	infocrop_cul_id	HD2285
&	REPLACE	wofost_cul_id	whtspa.dat
&	REPLACE	cropsyst_cul_id	wheat

Soil data. Some additional DOME functions are provided in the soil data. The ROOT_DIST() function allows the DSSAT array, soil root growth factor, to be set based on a value in the topsoil, the depth of the topsoil, and the depth at which the exponentially decaying function reaches 2% of the value in the topsoil. More information about this function is given here: http://research.agmip.org/display/itwiki/ROOT_DIST+-+Root+distribution+function.

The second function used here computes stable organic C as a function of soil depth based on the STABLEC() function, described here: http://research.agmip.org/display/itwiki/STABLEC+-+Stable+C+fraction+distribution+in+soil+layers.

The MULTIPLY() function simply takes one value as an argument, multiplies it by a constant, and returns the requested variable. In this case, the inert organic C array is computed as 0.9 times the stable organic C computed with the previous function.

!	Dome operator	Variable to be modified	Value or Function	Function arguments		
!		**Root distribution factor**	**Root distribution function**	**Value in topsoil (%)**	**Depth of topsoil (cm)**	**Depth at which SLRGF is 2% (cm)**
!	FILL	SLRGF	ROOT_DIST()	1	20	180
!		**Stable organic C array (%)**	**Distribution of stable organic C function**	**Fraction of stable organic C in topsoil (fraction)**	**Depth of topsoil (cm)**	**Depth at which SOM3 is ~98% of total (cm)**
&	FILL	SLSC	STABLEC()	0.55	20	60
!		**Inert organic C array**	**Multiply function**	**stable organic C**	**multiplier**	
&	FILL	SLIC	MULTIPLY()	$SLSC	0.9	

<u>Simulation controls</u> are model-specific and are never contained within the data collected in the field. These contain information such as when to start the simulation. In this example, the modeler chose to start the simulation 30 days prior to the planting date by using the OFFSET_DATE() function. The date associated with initial conditions was set to the start of simulation date.

!	Simulation Controls				
!	**Dome operator**	**Variable to be modified**	**Value or Function**	**Function arguments**	
!		**Start of simulation date**		**Planting date**	**minus 30 days**
&	REPLACE	SDAT	OFFSET_DATE()	$PDATE	-30
!		**Date for initial conditions**	**= Start of simulation date**		
&	FILL	ICDAT	$SDAT		

Additional examples of AgMIP formatted JSON and DOME data can be found on the AgMIP Github site:

https://github.com/agmip/json-translation-samples.

Connecting Experimental data with DOME data through Linkage files

If only one DOME file is specified, linkage is easy. The specified DOME is associated with each experiment and treatment (or survey element) in the dataset. But often, it is necessary to specify multiple DOMEs. For example, if the data set includes different soil types, then initial conditions may be specified for each soil type. In our Hot Serial Cereal example, no linkage file was used because the same field overlay DOME was applied to all treatments of the experiment. But if different DOMEs were applied, they would be specified in a linkage file (*.lnk) as shown below. In this sample, each EXNAME is associated with one or more field overlay files. The name of the DOME is constructed from the metadata block at the top of each DOME. Multiple DOMEs can be separated by the "|" (pipe) symbol, and they are applied in the order specified. Seasonal strategy DOMEs allow multiple years of execution from a single year of data and allow rules for automatic planting and irrigation to be specified for models that allow such rules. In this example, two field overlay DOMEs were specified and 3 seasonal strategy DOMEs.

DOME linkage specification:

#	EXNAME	FIELD_OVERLAY	SEASONAL_STRATEGY
*	NIORO001	NIORO------Soil1	NIORO-4----0XFX-CURRENT
*	NIORO002	NIORO------Soil1	NIORO-3----0XFX-CURRENT
*	NIORO013	NIORO------Soil2	NIORO-4----0XFX-CURRENT
*	NIORO017	NIORO------Soil2	NIORO-5----0XFX-CURRENT

When multiple DOME files are applied to a dataset, they should be "zipped" together and loaded into QuadUI as a zip archive file.