Integrating the Process-based Simulation Model DNDC into GIS

Stefan Huber¹, Georg Bareth¹ and Reiner Doluschitz¹

Abstract
Several trace gases, such as methane, nitrous oxide, nitric oxide, and ammonia, affect global climate change because of their radiative or chemical effects in the atmosphere. An important source of the four trace gases is soil. Agricultural land-use imposes a great deal of disturbance on soils by farming practices leading to elevated trace gas emissions. To assess emissions at regional scale we developed the software DNDC-GIS which integrates the mechanistic trace gas simulation model DNDC in a geographical information system (GIS). Additionally, to overcome the lengthy computation times, we implemented a load balancing module which allows deploying computational resources in a local-area network.

1. Introduction
There is a general agreement among scientists that several trace gases, such as methane (CH₄), nitrous oxide (N₂O), nitric oxide (NO), and ammonia (NH₃), affect global climate change because of their radiative or chemical effects in the atmosphere. An important source of the four trace gases is soil (Li 2000). Agricultural land-use imposes a great deal of disturbance on soils by farming practices including tillage, fertilization, irrigation, manure amendment, weeding, and liming. These anthropogenic activities elevate soil trace gas emissions and, hence, play an important role in the atmospheric balance of the trace gases. While field measurements have contributed extensively to our understanding in spatial and temporal variation of trace gas fluxes at site-scale, there is still a lack of methodologies to analyze emissions at regional scale. Paustian et al. (1997) propose a general framework for regional analyses of soil carbon which requires the integration of dynamic models, representing the feedbacks and interactions between soil processes, with information about the biotic and abiotic variables which drive these soil processes. Within a region, these driving variables vary spatially (and temporally) and therefore geographic information system (GIS) technologies can be employed to organize such information. We adapted this

¹ University of Hohenheim, Agricultural Informatics (410A), D-70593 Stuttgart, Germany, email: shuber@uni-hohenheim.de
framework for analyzing trace gas emissions from agricultural land-use by integrating the simulation model DNDC (Li 2000) into a geographical information system which we call DNDC-GIS.

The main disadvantage of this approach is its computational load, because the simulation model must be executed for every subset of input data. In some cases this may lead up to several thousand simulation runs for a study region. One approach to alleviate this disadvantage is to distribute the computational load over several computers in a local-area network increasing the throughput of computation (Foster 1995). Therefore we have extended DNDC-GIS with a module which enables us to distribute the workload over our department network.

This paper describes the software architecture of DNDC-GIS, the load balancing module, and presents results of a performance experiment by employing the latter.

2. Integration of DNDC into GIS

The software architecture of DNDC-GIS is shown in Fig. 1. The main components of the system are the commercial software ArcMap (ESRI 2001), a geodatabase, the DNDC model and the module RUN_SIM BUTTON which is responsible for configuration and transfer of data between the geodatabase and DNDC.

![Component model of DNDC-GIS](image)

Fig. 1: Component model of DNDC-GIS

ArcMap is part of ESRI’s ArcGIS, a scalable system of software for geographic data creation, management, integration, analysis and dissemination. ArcMap provides data display, query and analysis. ArcGIS is built on a technology framework known as ArcObjects, which is a collection of software components with GIS functionality and programmable interfaces. This technology is based on the COM (Mi-
The spatially distributed input variables for the DNDC model are stored in a MS Access 2000 geodatabase. As a test database we selected soil-land-use-climate information provided by earlier studies of a dairy region in southern Germany (Bareth et al. 2001).

The C++ source code of the DNDC model was modified and encapsulated in the in-proc COM-server object **DNDC_kernel_com** exposing the interface **IDNDC_SIM** in Microsoft Visual C++ 6.0. This interface is the only entry point to the simulation model and provides the method **run_DNDC(INPUT*, RESULT*)** which takes pointers to an **INPUT** and a **RESULT** data type as arguments. **INPUT** is a user defined data type encapsulating the necessary input data, such as soil, climate and agricultural management data. The results (e.g. yearly trace gas fluxes) of each simulation run are then stored in a **RESULT** structure.

As mentioned above ArcGIS allows extension of its base functionality by using ArcObjects. Therefore we have extended ArcMap by implementing the **ICommand** interface in the COM-object **RUN_SIM_BUTTON**. This is necessary to incorporate this module into ArcMap. **RUN_SIM_BUTTON** is responsible for configuration and data transfer between **DNDC_kernel_com** and the geodatabase. For configuration of data input and output it offers a graphical user interface.

### 2.2 Increasing performance with parallel computing

To increase performance, DNDC-GIS has an optional module which allows to distribute the computational load among several worker PCs in a network (Fig. 2). **RUN_SIM_BUTTON** obtains a reference to the **JOB_DISTRIBUTOR** module. The latter is a controller sending parcels of work (25 input data sets in one parcel) to **JOB_WORKER** modules on different computers in the network via the Transmission Control Protocol (TCP). **JOB_WORKER**, having a reference to **IDNDC_SIM**, unwraps the transferred parcel and passes the input data to **DNDC_kernel_com**. The returned results are stored in a buffer and sent back to **JOB_DISTRIBUTOR** as a parcel.

The network communication is based on the WinSock 2.0 API (Quinn/Shute 1998). WinSock offers asynchronous operation mode by calling the general-purpose API function **WSAAsyncSelect( )** which requests the WinSock DLL to notify the application of one or more events, such as connection completion and data arrival. Hence it allows an easy implementation of the message passing paradigm for parallel computing (Foster 1995).
3. Performance test

To test the optional load balancing module a small experiment was run on our 10 Mbit Ethernet department network. DNDC_kernel_com and JOB_WORKER were installed on eight desktop computers (PC2-PC9) each running Windows 2000. Processor speed and random access memory (RAM) benchmarks are shown in Table 1.

ArcMap, the geodatabase and RUN_SIM_BUTTON were installed on PC1 which acted as a work distributor via the JOB_DISTRIBUTOR module. Four different configurations were tested with 10 replication runs on 5000 input datasets over a simulation period of two years: O1. Local mode (e.g. DNDC-GIS without optional load balancing module solely on PC1); O2. Small cluster mode (e.g. PC1 as work distributor and PC2 to PC4 as workers); O3. Medium cluster mode (e.g. same as O2 but with PC2 to PC6 as workers); O4. Big cluster mode (e.g. same as O2 but with PC2 to PC9 as workers).
Table 1: Benchmarks of controller (PC1) and workers (PC2 to PC9). O1 is standard DNDC-GIS local operation mode without load balancing. In O2 to O4 work is distributed to the computers as denoted in the table by •.

<table>
<thead>
<tr>
<th>PC-ID</th>
<th>Processor, RAM (MB)</th>
<th>Operation Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>Athlon XP 1700, 512</td>
<td>O1  O2 O3 O4</td>
</tr>
<tr>
<td>PC2</td>
<td>2x PIII 1000, 512</td>
<td>• (•) (•) (•)</td>
</tr>
<tr>
<td>PC3</td>
<td>Athlon XP 1500, 768</td>
<td>• • • •</td>
</tr>
<tr>
<td>PC4</td>
<td>PIII 1400, 265</td>
<td>• • • •</td>
</tr>
<tr>
<td>PC5</td>
<td>Athlon XP 1900, 512</td>
<td>• • • •</td>
</tr>
<tr>
<td>PC6</td>
<td>PIII 700, 256</td>
<td>• • • •</td>
</tr>
<tr>
<td>PC7</td>
<td>Athlon XP 1000, 256</td>
<td>• • • •</td>
</tr>
<tr>
<td>PC8</td>
<td>PIII 500, 256</td>
<td>• • • •</td>
</tr>
<tr>
<td>PC9</td>
<td>PIII 1000, 512</td>
<td>• • • •</td>
</tr>
</tbody>
</table>

Fig. 3 shows the average throughput, defined as the amount of computed input datasets per minute, of the different operation modes.

![Throughput graph]

Fig. 3: Throughput of processed input datasets (N) per minute for different operation modes of DNDC-GIS.

The results clearly show that distributing the workload over the network increases the throughput of computed datasets. The performance gain depends on the number of worker PCs. Eight worker PCs, as compared to the local mode O1, increased the throughput by a factor of 5.
4. Conclusions

In this contribution, we described a software architecture for the integration of an agro-ecosystem model into GIS for regional analysis of trace gas emissions in meso scale. The integration of DNDC into GIS enables the fully automatic spatial modeling of the contribution of agricultural land use on the atmospheric balance of the trace gases. All input data for DNDC derive from the geodatabase.

One bottleneck of regional analysis is the computational load. We proposed a simple implementation of the message passing paradigm of parallel computing. This enabled us to use the computational power of several computers in a network forming a virtual super computer.

Bibliography


ESRI (2001): What is ArcGIS, Environmental Research Institute, Redlands


