PSAS Report

Goals for the first semester of 2005

The goals for the first semester of 2005 were:

1. Investigate the reasons for the lack of binary reproducibility when changing the number of processes; if possible, correct them.
2. Vectorize and optimize both parts of PSAS, on sequential runs for the SX6.

The algorithm of the parallel version of PSAS comprises five steps:

- Partition observations using inertial recursive bisection scheme;
- Matrix block distribution of the innovation matrix;
- Solve the innovation equation using the Conjugate Gradient method;
- Partition of the specified forecast error covariance matrix and the generalized interpolation from the analysis grid to the observations;
- Solve the analysis equation.

Reproducibility

There are two aspects of reproducibility for PSAS. First, the partial vector sum algorithm in the matrix block distribution and analysis increment where, for different numbers of precessors, additions may be performed in a different order that cannot guarantee the bitwise identical results. For PSAS this can only be guaranteed for runs on the same numbers of precessors.

The second aspect of reproducibility is that results may vary up to the middle order bits when the number of regions is changed. The number of regions is a special case because it is coupled to the configuration of the parallel computer. This parameter must not only be a power of two, because of the recursive bisection algorithm, but it must be an integer multiple of the number of processors in order that the analysis equation be load balanced. Another problem with trucation may arise for inhomogeneous observations patterns. The parallel partitioner will generate relatively large regions where data are sparse. In this case the use of the data centroids of regions as a criterion for applying the correlation cutoff may be inconsistent. A more rigorous approch may base the cutoff on the minimum distance between vertexes of regions. This in turn will adversely affect load balance of analysis equation, which assumes blocks of the forecast error covariance matrix to be approximately of the same size.

Optimization

The performance evaluation shows that the part of PSAS, responsible for the data ingestion,

preparation and write out the analysis (*ana.x*), had very poor performance. It runs sequentially at 6 MFlops, 18% vectorization and vector lengths of 67 and it demands about 1000 seconds of execution. The second part of PSAS (the analysis itself - solve.x) runs sequentially at 832 MFlops, with 94.2% vector operation ratio and average vector length of 151, consuming 1866 seconds. The Table 1 shows the program information for this running. After the modifications in some routines in both parts of PSAS, we enhance the perfomance in PSAS. The program *ana.x*, running sequentially, now has 40 Mflops,

Table 1 - Program information without optimization for the two parts of PSAS

solve.x		ana.x	
****** Program Information ******		****** Program Information ******	
Real Time (sec)	: 1866.603702	Real Time (sec)	: 2866.424379 (1866.60 + 999.82)
User Time (sec)	: 1861.536445	User Time (sec)	: 973.152712
Sys Time (sec)	: 1.806679	Sys Time (sec)	: 7.970844
Vector Time (sec)	: 767.130805	Vector Time (sec)	: 156.679668
Inst. Count	: 284429187428	Inst. Count	: 178740435706
V. Inst. Count	: 26813532624	V. Inst. Count	: 617511157
V. Element Count	: 4073502861701	V. Element Count	: 41762955130
FLOP Count	: 1548813617560	FLOP Count	: 6024873024
MOPS	: 2326.636434	MOPS	: 225.952080
MFLOPS	: 832.008216	MFLOPS	: 6.191087
VLEN	: 151.919664	VLEN	: 67.631094
V. Op. Ratio (%)	: 94.051983	V. Op. Ratio (%)	: 18.993014
Memory Size (MB)	: 496.031250	Memory Size (MB)	: 304.031250
MIPS	: 152.792704	MIPS	: 183.671518
I-Cache (sec)	: 1.752371	I-Cache (sec)	: 2.137298
O-Cache (sec)	: 88.582894	O-Cache (sec)	: 92.417792
Bank (sec)	: 48.430037	Bank (sec)	: 74.605062

Table 2 - Program information with optimization for the two parts of PSAS

solve.x		ana.x	
****** Program Information ******		****** Program Information ******	
Real Time (sec)	: 1700.702570	Real Time (sec)	: 1929.151294
User Time (sec)	: 1689.551166	User Time (sec)	: 180.632316
Sys Time (sec)	: 2.256522	Sys Time (sec)	: 10.722667
Vector Time (sec)	: 727.347024	Vector Time (sec)	: 9.944475
Inst. Count	: 251429994628	Inst. Count	: 46789861848
V. Inst. Count	: 26616226107	V. Inst. Count	: 88607792
V. Element Count	: 4062174209977	V. Element Count	: 6194023724
FLOP Count	: 1548813632051	FLOP Count	: 7397160687
MOPS	: 2537.353153	MOPS	: 292.833968
MFLOPS	: 916.701230	MFLOPS	: 40.951480
VLEN	: 152.620217	VLEN	: 69.903827
V. Op. Ratio (%)	: 94.755904	V. Op. Ratio (%)	: 11.709975
Memory Size (MB)	: 496.031250	Memory Size (MB)	: 304.031250
MIPS	: 148.814667	MIPS	: 259.033726
I-Cache (sec)	: 3.255382	I-Cache (sec)	: 1.250926
O-Cache (sec)	: 79.969915	O-Cache (sec)	: 23.389751
Bank (sec)	: 30.073283	Bank (sec)	: 4.265875

Goals for the second semester 2005

The goals for the second semester were concentrated to adapt the global version of PSAS to regional version. The adaptation included changes in files structures and compilation. The second part of the goals were the performance analysis in the run time. The performance analysis showed two function (dist, alfa) with very big frequency spending a lager time, about 90% of total run time. Using the procedure to expand in line the functions, the performance became better and the run time was reduced to 30% of the time. The new performance analysis showed a subroutine (mindis) spending a larger time, about 80 % of total run time, low vector operatio ratio, small vector lengh and few vector time. This indicated the target to became the performance better. This part should be finish in the end of june 2006.