

1 **GSI-BASED 3DENSVAR DATA ASSIMILATION FOR THE**
2 **BAM-CPTEC/INPE: SINGLE LOW RESOLUTION**
3 **EXPERIMENTS**

4 CARLOS FREDERICO BASTARZ *†, DIRCEU LUIS HERDIES†

†*Center for Weather Forecast and Climate Studies, CPTEC/INPE, Cachoeira Paulista, SP, Brazil*

*Corresponding author address: Carlos Frederico Bastarz, CPTEC/INPE, Rod. Presidente Dutra, km 40, Cachoeira Paulista, SP, Brazil.

E-mail: carlos.bastarz@inpe.br

†Affiliation: Center for Weather Forecast and Climate Studies, CPTEC/INPE, Cachoeira Paulista, SP, Brazil; Post-Graduation Program in Meteorology at INPE, São José dos Campos, SP, Brazil

ABSTRACT

This paper aims to present the first results obtained with a GSI-based Hybrid Ensemble-Variational system applied to the Brazilian Atmospheric Model from Center for Weather Forecasts and Climate Studies at the National Institute for Space Research (BAM-CPTEC/INPE). The hybrid data assimilation system implemented is a 3DEnsVar, using the Ensemble Kalman Filter (EnKF) and the Gridpoint Statistical Interpolation (GSI) 3D-Variational analysis (3DVar) system. This is an upgrade to the previous GSI 3DVar system being used at CPTEC operations for Numerical Weather Predictions (NWP). The implementation was made using a previous established model and assimilation framework. A new background error covariance matrix was calculated, in order to properly account the ensemble covariance contribution to the static climatological part. An analysis and an evaluation of the updated system is presented and the first results for analyses and 5-day forecasts at TQ0062L028 model resolution are presented as well. The tested 3DEnsVar system at CPTEC shows an improvement for the BAM-CPTEC/INPE skill mainly over South America and Tropical region and endorses the use of ensemble covariances as a complement to the static part of the covariances as a way to reduce the misrepresentation of the background error covariances. Precipitation has been also evaluated in a comparison with observed data. Exercising with the hybrid system have shown an improvement over the pure 3DVar system in both the intensity and the distribution of the precipitation over the time mean. The results obtained with the implementation are encouraging and further improvements must be made in order to fully access the analysis and forecast skill, specially under higher resolutions.

²⁶ 1. Introduction

²⁷ Hybrid data assimilation systems using an ensemble Kalman Filter and a variational
²⁸ technique have been developed (e.g., Hamill and Snyder 2000; Lorenc 2003; Zupanski 2005)
²⁹ and applied focusing the sampling and the representation of the spatial-temporal variations
³⁰ of background error covariances, introducing the day-to-day statistics variations of the
³¹ background flow (the so called “errors of the day” - Corazza et al. 2003) to the static part
³² of the variational covariances. The representation of background error covariances is one of
³³ the main issues of the operational data assimilation and its specification for deterministic
³⁴ systems (e.g., variational systems) is of a primordial importance.

³⁵ Variational systems such as 3DVar uses a static background error covariance matrix,
³⁶ which means that its covariances are stationary in time and does not vary, thus the same
³⁷ error statistics are used for the whole assimilation process. A similar situation happens
³⁸ to a 4DVar system, which - in general, uses a pre-computed covariance background error
³⁹ matrix at the beginning of each assimilation window, which is at least propagated in time
⁴⁰ during the assimilation of observations using the tangent linear version of the forecast model.
⁴¹ Differences in the way that background error statistics varies with time due to the variations
⁴² in the background flow, leads to what we call flow-dependence. Anisotropy is related to the
⁴³ geometrical aspect of the correlation functions that made then more or less suited to the
⁴⁴ gradients represented in the background flow (e.g., like in a front line). Inhomogeneity is
⁴⁵ related to the spatial distribution of covariances at each gridpoint, and how the correlation
⁴⁶ functions are defined. These desirable features are attained when the covariances structures
⁴⁷ being represented are modeled with the use of operators which are able to describe physical
⁴⁸ processes within a dynamical nature.

⁴⁹ Flow-dependent covariances are a key feature of modern data assimilation systems, and its
⁵⁰ specification either at the beginning or during the assimilation window, allows the analysis to
⁵¹ correct account for the time and spatial dependency of the observation innovations being used
⁵² into the analysis increments, in both horizontal and vertical directions. Furthermore, desirable

53 features of the background error covariances also include anisotropy and inhomogeneity (non-
54 uniform statistics - Rabier 2005). Flow dependence stands for the background dynamics
55 and its spatial changes with time, which must be represented inside the background error
56 covariance matrix. Anisotropy and inhomogeneity are characteristics of the spatial distribution
57 of the covariances and are governed through correlation functions. Correlation functions are
58 modeled in such a way to allow the background error statistics to be non-uniform, i.e., to be
59 different among gridpoints.

60 Several efforts have been made in order to add some degree of anisotropy and flow-
61 dependence to static covariances. The development of the variational techniques for op-
62 erational use during the 1990's, include some sort of implicit modifications to the data
63 assimilation framework. Desroziers (1997) introduced a coordinate transformation to allow
64 the data assimilation system to better account for structures of frontal systems. This led to
65 an improved analysis with flow dependent covariances and anisotropic correlations.

66 The gradients represented in the background flow, also include important information that
67 could be used to allow anisotropy. Spatial background gradients can also be used to model
68 error statistics. Riishøjgaard (1998) showed how to use the humidity field to model correlation
69 functions that can be stretched to accommodate the covariances according to the gradients
70 represented in the background flow. Other approaches to model correlation functions that
71 lead to flow-dependent background error statistics, involve the use of wavelets with spectral
72 methods (e.g., Fisher 2003), but difficulties arise with the specification of these wavelets on
73 the sphere. Recursive filters are also another known method to derive quasi-Gaussian shapes
74 to use in background error covariance modeling. This approach has been used successfully in
75 several applications; e.g., Hayden and Purser (1995) applied isotropic recursive filters to the
76 processing of the National Environmental Satellite and Data Information Service (NESDIS);
77 more recently, recursive filters have been applied to the Gridpoint Statistical Interpolation
78 (GSI) to model the application of background error covariances (Wu et al. 2002).

79 Within these efforts, other approaches have been developed and are more related to the

80 representation of nonlinearities of the background flow into the background error statistics.
81 The 4DVar (Thépaut and Courtier 1991) was developed as an extension of 3DVar (Lorenc
82 1986) and can implicit evolve covariances inside the assimilation window. Corazza et al. (2003)
83 added flow-dependence to covariances by accounting for the “errors of the day”, using bred-
84 vectors. Hybrid methods (e.g., Hamill and Snyder 2000) were introduced taking advantage of
85 the ensemble Kalman Filter methods. This new approach has several advantages including
86 the fact that operational centers have already experienced the variational technique and have
87 some experience with ensemble Kalman Filters. Each of these methods can be taken as a
88 compliment to ameliorate the deficiencies of the other (Wang et al. 2007, 2009). Recently,
89 hybrid methods with real applications (e.g., Wang et al. 2013; Clayton et al. 2013) using
90 covariances estimated from a ensemble Kalman Filter, are linearly combined to the 3DVar
91 static covariances matrix.

92 Most recent developments at CPTEC have focused on its independence on observations
93 pre-processing, quality control and analysis. In the past few years, some analysis systems
94 have been tested at CPTEC, including the GSI. GSI have been used in several operational
95 NWP centers and is the data assimilation chosen for regional and global applications at the
96 center. CPTEC has plans to use GSI to provide analysis for some of its regional models
97 and the center is making real progress with its new general circulation model (the Brazilian
98 Atmospheric Model - BAM-CPTEC/INPE; Figueroa et al. 2016).

99 With the objective of provide the BAM-CPTEC/INPE model with its own analysis in a
100 data assimilation cycle, in this paper are described the details of the implemented system
101 and the first results obtained with the GSI-based hybrid 3DEnsVar system.

102 This paper is organized as follows: Section 2 describes the strategy and the details of the
103 implemented hybrid system and the static part of the background error covariance matrix;
104 Section 3 presents a description of the idealized experiments to test the hybrid system. Results
105 from the experiments carried out are represented in Section 4 and finally, Section 5 presents
106 the discussion regarding the results and future plans.

¹⁰⁷ 2. CPTEC Global Hybrid 3DEnsVar System

¹⁰⁸ In the past years, hybrid data assimilation systems have shown an increasing amount of
¹⁰⁹ development and several centers have been made efforts to take advantage of its benefits.
¹¹⁰ Hybrids systems using the 3DVar and a ensemble Kalman Filter (EnKF) appears to be
¹¹¹ common choice to start with. The GSI system comprises a collection of routines that are
¹¹² already prepared to run such a hybrid system using the EnKF to provide the ensemble
¹¹³ covariances to compliment the static covariances used in 3DVar.

¹¹⁴ Since 2012 CPTEC have been experimenting the GSI 3DVar analysis with its global
¹¹⁵ General Circulation Model (GCM) for NWP until 7 days at a TQ0299L064 spectral model
¹¹⁶ resolution (roughly, 45 km at the Equator, with a corresponding grid of 450 X 900 X 64
¹¹⁷ points, *lat X lon X lev*). This version of the system was performed with a GSI's example
¹¹⁸ global background error covariance matrix (with a grid of 386 X 768 X 64 points, *lat X lon*
¹¹⁹ X *lev*). Although this configuration seems to be not the ideal, it was a first attempt to run a
¹²⁰ global analysis with radiance data assimilation at a higher resolution at CPTEC.

¹²¹ With the public release of the hybrid version of GSI in 2015, CPTEC started to plan
¹²² a revamped version of its global data assimilation system. A new version of the BAM-
¹²³ CPTEC/INPE model was released in the summer of 2015 and was planned a test of running
¹²⁴ the BAM-CPTEC/INPE with its own analysis, but using a proper background error covariance
¹²⁵ matrix and also testing with the new hybrid system structure of GSI.

¹²⁶ Exercise with the hybrid ensemble-variational assimilation technique with the BAM-
¹²⁷ CPTEC/INPE, was made using the available ensemble algorithms. Four experiments were
¹²⁸ carried out: one experiment using the Ensemble Square Root Filter (EnSRF - Whitaker and
¹²⁹ Hammill 2002) and a second experiment with the Ensemble Kalman Filter (EnKF - Evensen
¹³⁰ 1994) were designed to test the 3DEnsVar hybrid system. Another experiment with a pure
¹³¹ 3DVar using the BAM-CPTEC/INPE in a cycled data assimilation and the realization of
¹³² the BAM-CPTEC/INPE with the National Centers for Environmental Predictions (NCEP)
¹³³ analysis were also performed. For the three experiments using data assimilation, the new

134 static version of the background error covariance matrix were used. As we have made a new
135 background error covariance matrix for use in the experiments with the traditional 3DVar
136 and the hybrid data assimilation, Subsection “a” shows the details and main aspects the the
137 new static matrix. Subsections “b” and “c” are dedicated to present the BAM-CPTEC/INPE
138 model information and the procedures and the main setup for the hybrid 3DEnsVar system,
139 respectively. A detailed description of the experiments is given in Section 3.

140 *a. Static Background Error Covariance Matrix*

141 The background error covariance matrix (**B**) has the role to filter and propagate spatially
142 the observation information (Berre et al. 2013), contributing to transform part of the
143 observations innovations into analysis increments.

144 The calculation of the background error covariance matrix for the GSI system used in
145 this work, involved the use of 730 pairs of 48 and 24 hours forecasts, valid for the 0000 and
146 1200UTC, distributed evenly over 1 year (2013). The resolution of the forecasts matches
147 the resolution of the forecast model used in the experiments, i.e., TQ0062L028 (with 192
148 x 96 x 28 - *lat X lon X lev* grid points). The forecasts pairs were organized as follows:
149 considering 2014010100, the first valid forecast pair was generated with the analysis from
150 2013123100 (valid for a 24 hours forecast) and 2013123000 (valid for a 48 hours forecast).

151 The methodology used to calculate the amplitudes (i.e., the variances and covariances) is
152 based on the NMC method (Parrish and Derber 1992). This method claims that spatial
153 correlations of the model errors are similar to the spatial correlations of the differences
154 between the 48 and 24 hours forecasts. The NMC method was chosen for the background
155 error covariance calculations due to the convenience of having previously analyses made from
156 operational routines. These analyses were used to integrate the BAM-CPTEC/INPE model
157 at TQ0062L028 resolution for one year. Within the algorithm involved in the calculations,
158 there is also a procedures to remove the bias in the vertical model column, which were also
159 used. Covariances were calculated for the standard GSI control variables, i.e., streamfunction

160 (ψ) , unbalanced velocity potential (χ), unbalanced temperature (T), unbalanced surface
161 pressure (ps) and the normalized relative humidity (rh). The sea surface temperature (sst)
162 is the same as the one used on the GFS-NCEP background error covariance matrix (i.e., the
163 Real-Time, Global, Sea Surface Temperature - RTG_SST analysis, from which the variances
164 and length scales were interpolated to the target grid).

165 In practice, although GSI doesn't explicit construct \mathbf{B} , it applies the covariances of \mathbf{B}
166 using recursive filters (Purser et al. 2003a,b):

$$\mathbf{B} = \mathbf{B}_z(V^1 \mathbf{B}_x^1 \mathbf{B}_y^1 \mathbf{B}_x^1 V^1 + V^2 \mathbf{B}_x^2 \mathbf{B}_y^2 \mathbf{B}_x^2 V^2) \mathbf{B}_z \quad (1)$$

167 where,

- 168 • V^1 and V^2 are the standard deviations of each control variable (from the \mathbf{B} file);
- 169 • \mathbf{B}_x , \mathbf{B}_y and \mathbf{B}_z are the recursive filter applications in the directions of x (west-east), y
170 (south-north) and z (vertical);
- 171 • \mathbf{B}^1 and \mathbf{B}^2 represents the application of the recursive filter in the horizontal scales (x
172 and y).

173 An important property of the recursive filters in the GSI framework, is the adjustment of
174 the amplitudes and the length scales in order to make the aspect of the covariances anisotropic,
175 letting them to adjust to the background flow. These parameters and scales used in the
176 application of the recursive filters are sort of empirical and a discussion of it can be found in
177 Wu et al. (2002). The resulting amplitudes of the calculated \mathbf{B} are presented in Section a.

178 b. *The Brazilian Atmospheric Model (BAM)*

179 CPTEC has been made efforts to continuously develop and improve its own GCM for
180 operational use at the center. As a result of this efforts, the Brazilian Atmospheric Model
181 (BAM-CPTEC/INPE, Figueiroa et al. 2016) brings the most recent advances in both model

182 dynamics and physics. Although the version of the BAM-CPTEC/INPE model used in this
183 work is not exactly the same version used in Figueroa et al. (2016), the version used in the
184 experiments will be referred as “BAM v0.0”, which includes the same dynamical core and
185 a simplified physics package (mostly described in Cavalcanti et al. 2002). This simplified
186 physics package is based on a large scale condensation due to the chosen model resolution
187 (TQ0062L028, roughly 200 km at the Equator) and the number of ensemble members (40
188 plus 1 of the hybrid analysis). Furthermore, microphysics was avoided due to the intense
189 computational requirements (Silvio N. Figueroa, personal communication).

190 The BAM-CPTEC/INPE is a spectral model representing mass and winds in terms
191 of vorticity and divergence, with a pure sigma vertical coordinate, used at the center for
192 operational NWP (current resolution is TQ0666L064, roughly 20 km at the Equator) up
193 to 5 days, and - at a coarse resolution (TQ0126L028, roughly 100 km near the Equator)
194 for extended NWP up to 15 days, using a ensemble prediction system based on Empirical
195 Orthogonal Functions (EOF) for the perturbation of the initial condition. Seasonal and
196 climate simulations are made as well. In every case, except for the experiments related here,
197 analysis are taken from GFS-NCEP.

198 *c. Hybrid Ensemble-Variational Analysis Cycle*

199 The 3DEnsVar is the hybrid ensemble-variational system intended to be the next generation
200 of atmospheric analysis at CPTEC. This system was exercised in an experimental basis at
201 CPTEC for weather forecasts only (up to 5 days at TQ0062L028 model resolution). Plans
202 include to use the ensemble of analysis from the hybrid system to replace or to augment
203 the current size (15 members) and consequently the spread of the center global Ensemble
204 Prediction System (EPS, described in Mendonça and Bonatti 2009; Cunningham et al. 2015),
205 for extended atmospheric forecasts (up to 15 days, at TQ0126L028 model resolution - current).
206 The hybrid system used in this work is coupled with the BAMv0 model through an interface
207 that reads the spectral forecasts from the BAMv0 model and rewrites the spectral coefficients

208 for use as a background for the GSI system. This interface is currently being updated so in
209 the near future there will be not needed anymore (in fact, the CPTEC staff is updating GSI
210 in order to read the BAMv0 background files as a native model). In the current setup, the
211 interface writes an intermediate file that mimics that one from GFS-NCEP. When the GSI
212 analysis is completed, a program is executed to rewrite the file with a proper header and
213 array order.

214 The current hybrid setup requires a realization of the GSI observer using the mean
215 ensemble as a background. No analysis is made at this point and this step is only necessary to
216 write the diagnostics files from the departures (i.e., $y^o - \mathbf{H}x_k^b$). The next step is the realization
217 of the observer using the departures related to the ensemble mean, but using every ensemble
218 member as a background. In the setup tested, an ensemble of 40 members was used. This
219 number represents a reasonable ensemble size, which fits all the computational requirements,
220 regarding the allowed computational time and disk space without compromise the efficiency.
221 Once the observer ensemble of departures is computed within GSI, the EnKF/EnSRF is
222 applied to update the ensemble of backgrounds. In this step, the ensemble of analysis are
223 stored in disk and used to initialize a new ensemble of BAMv0 forecasts valid for the next
224 analysis cycle. The final deterministic hybrid analysis is then achieved using the covariances
225 from the updated ensemble of backgrounds, which are blended with the static pre-computed
226 background error covariance matrix. The resulting analysis is then used to initialize the
227 BAMv0 model, from which 5-days forecasts are performed. Figure 1 shows a schematic
228 diagram of the analysis cycle, with some of the main equations used at each step. In order to
229 make it more clear, a time dimension was added to the terms in the equations indicating the
230 time position of the cycle (e.g., time= $t - 1, t, t + 1$ and so on).

231 The hybrid 3DEnsVar system minimizes a 3D variational cost function in the same fashion
232 as 3DVar. The main difference is regarded to the way the background error covariance matrix
233 is defined and applied. Equation 2 (presented here in a more general way) represents the
234 3DEnsVar cost function, where x' is the new analysis increment.

$$J(x') = \frac{1}{2}(x')^T(\alpha_1\mathbf{B} + \alpha_2\mathbf{P}^b)^{-1}(x') + \frac{1}{2}[y_o - \mathbf{H}(x')]^T\mathbf{R}^{-1}[y_o - \mathbf{H}(x')] \quad (2)$$

235 Note that the term $(\alpha_1\mathbf{B} + \alpha_2\mathbf{P}^b)$ represents the linear combination between the pre-
 236 computed static \mathbf{B} and the ensemble flow-dependend ensemble covariances \mathbf{P}^b . The coefficients
 237 α_1 and α_2 are the weights assigned to each parcel of the linear combination, trough which we
 238 control the amount of contribution of the ensemble covariances to the static part.

239 In Equation 2, x' is defined as:

$$x' = x + \sum_{k=1}^K (a_k \circ x_k^e) \quad (3)$$

240 This analysis increment is defined as the sum between the standard analysis increment
 241 (x) due to the static background error covariance contribution and the summing between
 242 the extension of the control variable (a_k , to account for the ensemble localization) and the
 243 ensemble perturbation of each member ($x_k^e = \frac{(x_k^b - \bar{x}^b)}{\sqrt{K-1}}$). A more detailed explanation of the GSI
 244 methodology to extend the variational framework to accommodate the ensemble covariances
 245 can be found in Wang et al. (2007, 2008a,b), Wang (2010) and Wang et al. (2013).

246 To test the 3DEnVar hybrid system, the ensemble Kalman Filter (EnKF - Evensen 1994)
 247 and the ensemble Square-Root Filter (EnSRF - Whitaker and Hamill 2002) were tested
 248 and exercised to test the implementation.

249 Despite the fact that the 3DEnsVar analysis cycle have been established for the BAMv0
 250 model, some caveats have been found and will demand attention on future revisions of this
 251 system. The system works in a single resolution, i.e., both the control and the ensemble of
 252 background/analysis are run at the same model resolution (currently TQ0062L028). As a
 253 consequence, a recentering of the ensemble mean around the hybrid analysis, is not made.
 254 Another point to mention is the fact that the bias mass and angle are not being updated,
 255 i.e., during the experiments, no mass or angle of the radiance channels and instruments were
 256 updated. This decision was taken because, as a testing system, no compromise was made to
 257 keep the radiance information error updated. Also, there is no vertical localization within

258 the EnKF/EnSRF, so a fixed value is used for the entire vertical column.

259 3. Experimental design

260 To test the 3DEnsVar system, a set of experiments were idealized in order to verify the
261 analysis and forecasts skill (in terms of the Root Mean Square Error - RMSE and the Anomaly
262 Correlation - AC), by comparing them with a pure (control) 3DVar analysis. The pure 3DVar
263 analysis, in turn, was compared to an offline analysis experiment by integrating the BAMv0
264 model with an independent analysis. The next subsection are dedicated to explain the design
265 of each experiment.

266 a. Control Experiment

267 The control experiment is intended to be a 3DVar experiment run with the new background
268 error covariance matrix. This experiment is the control run and is compared to the hybrid
269 experiments (i.e., using the EnKF and the EnSRF algorithms). The control was ran using
270 the variational analysis in TQ0062L028 resolution and the model configuration (both the
271 dynamics and the physics) is the same among the hybrid experiments. The analysis cycle
272 is setup in the following manner. A set of three background forecasts of 3, 6 and 9 hours
273 (where the central analysis is always at 6 hours) are used by GSI for the update with the
274 observations in a First Guess at Appropriate Time (FGAT) approach. Observations include
275 conventional (u , v , T , q and ps) from several sources (e.g., the Global Telecommunication
276 System - GTS) and unconventional data, like radiance, retrievals and Global Positioning
277 System - Radio Occultation (GPS-RO).

278 Table 2 from Section c summarizes the types of observations included in the experiments.

279 The same set of observations are used for the hybrid experiments as well.

280 The analysis cycles continues when GSI writes the analysis file for the atmosphere and
281 the surface. It is worth to note that only the atmospheric part of the analysis was used to

282 initialize the BAMv0, although the model restart files were used to initialize the surface fields,
283 as well clouds and radiation. This same approach is taken for the hybrid experiments.

284 Another idealized experiment is the offline analysis (i.e., without an analysis cycle). In
285 this experiment the BAMv0 where integrated in time with a hi-resolution NCEP spectral
286 operational analysis which was chopped at each analysis time (i.e., 00, 06, 12 and 18UTC).
287 From this experiment, 120 hours forecasts are generated, in the same fashion as the 3DVar
288 run. This offline experiment is intended to serve as a control for the pure 3DVar experiment,
289 whereas the 3DVar analysis experiment serves as a control for the hybrid experiments.

290 *b. Hybrid Experiments*

291 Exercising with the 3DEnsVar was made cycling the system with the BAMv0 model
292 between December 2012 and January 2013. A spin up time was considered and the first
293 month of the simulations was taken apart from the results. One of the reasons for this
294 practice is the fact that the initial ensemble of analysis was generated by running the BAMv0
295 model from a single deterministic analysis and forecasts made up to 30 days at each 12 hours
296 (i.e., twice a day). From this, a time series of 60 forecasts ranging from 12 hours up to 720
297 hours from the analysis time. The first 40 forecasts were kept and carefully modified to reflect
298 the background time for the first analysis to be generated in all 3DEnsVar experiments.

299 To assess the behaviour of the 3DEnsVar system using the BAMv0 model as a background
300 and its new background error covariance matrix, four cycled data assimilation experiments
301 were ran. The experiment called 3DVar represents the variational GSI/3DVar experiment with
302 the new **B** (considering $\alpha_1 = 1$, i.e., no ensemble covariance contribution). The experiment
303 called EnKF50 represent the same cycle experiment ran using the hybrid **B** through the use
304 of the EnKF algorithm (considering $\alpha_2 = 0.5$, i.e., considering 50% of contribution of the
305 ensemble covariances). Experiment called EnSRF50 is the same experiment as EnKF, but
306 using the EnSRF, where the observations are not perturbed. The experiments EnKF and
307 EnSRF were run with an ensemble size of 40 members. In addition to experiments EnKF50

308 and EnSRF50, the ensemble contribution to the static part of the covariances were also tested
309 with 75% of contribution. Experiments using the EnKF and EnSRF for this relative amount
310 of ensemble contribution are called EnKF75 and EnSRF75, respectively.

311 Table 1 summarizes the experiments designed to exercise the 3DEnsVar system, ranking
312 the ensemble covariance contribution for each experiment.

313 *c. Observational Data*

314 The observational data used for the experiments with a cycled analysis using the 3DVar GSI
315 and the EnKF/EnSRF algorithms, are the same. Table 2 for a summary of the observational
316 data used.

317 *d. Model Setup*

318 In all experiments, regardless the analysis used, the same model configuration had been
319 taken. Table 3 summarizes the main options adjusted within the dynamical core and physics
320 package. It is worth to note that in the data assimilation cycle (regardless the hybrid or
321 the pure variational option), the atmospheric model was restarted at every cycle (00, 06,
322 12 and 18UTC). The BAMv0 restart files comprises the previous states of the land-surface,
323 clouds and radiation. Sea surface temperature, snow cover and soil moisture are read from
324 separated files and the dynamical state (i.e., momentum) are provided by the atmospheric
325 analysis. These conditions are valid for the generation of the background files (3h, 6 and 9
326 hours forecasts) used by the GSI.

327 *e. GSI Setup*

328 This subsection is intended to give an overview of the basic and main GSI options that
329 was chosen to complete the experimental design. GSI has many different option that can
330 be adjusted in order to allow the system run with specific procedures or to tune the system

331 through the use of knobs that must be setup empirically. All of these option can be found
332 on the official GSI manual, available at the Development Testbed Center (DTC) website.
333 Three of these general option worth to mention here (outer and inner loops, choice for
334 humidity control variable and the moisture constraint), as they were kept the same among
335 all experiments using GSI.

336 OUTER AND INNER LOOPS

337 In all experiments, GSI was configured to run with 3 outer loops. The first outer loop was
338 set with a maximum number of 100 inner iterations. The second outer loop was set with 150
339 inner iterations and the third outer loop, with 50 iterations. The options for the outer and
340 inner loops was chosen to be fairly large to allow for the convergence of the cost function.
341 For these initial experiments with the hybrid systems, no compromise was made to exercise
342 the system with performance in mind.

343 HUMIDITY CONTROL VARIABLE

344 GSI has two option to treat the humidity control variable. They are the pseudo relative-
345 humidity and the normalized relative humidity. As we were testing a new background error
346 covariance matrix, we decided to choose the normalized relative humidity as the humidity
347 control variable in all experiments. The normalized relative humidity allows the relative
348 humidity to change in the inner loops in accordance to changes in the surface pressure,
349 temperature or specific humidity.

350 MOISTURE CONSTRAINT

351 Supersaturated and negative moisture are non-physical solutions that arises from computa-
352 tion modes. In GSI, there is an option that can be tuned in order to control the supersaturated
353 and negative moisture. Several tests were made in order to achieve reasonable values to

354 allow the stability of the system. The values found are 5.0 for the supersaturated moisture
355 constraint and 0.005 for the negative moisture constraint.

356 4. Results

357 This section shows the results obtained with the application of a hybrid background error
358 covariance matrix through the use of a 3DEnsVar system for the BAMv0 model. Combined
359 with the experimental results, the main characteristics of the calculated background error
360 covariance matrix are shown. Results from the experiments with the hybrid 3DEnsVar
361 are divided into four categories: 1) the ensemble innovation statistics; 2) analysis skill
362 (taken in terms of RMSE); 3) forecasts skill (up to 120 hours, taken in terms of AC) and
363 4) precipitation assessment (taken in term of a comparison with the GPCP data). The
364 evaluations taken in terms of RMSE and AC, are made to the regions Global (GL), North
365 Hemisphere (NH - lons:0°-360°; lats:20°N-80°N), Tropics (TR - lons:0°-360°; lats:20°S-20°N),
366 South Hemisphere (SH - lons:0°-360°; lats:80°S-20°S) and South America (SA - lons:0°-360°;
367 lats:49.875°S-11.375°N). The anomaly correlations are presented along with t-Students tests
368 to verify the significance of the results.

369 a. Standard Deviations of the Static **B**

370 Experiments with a cycled data assimilation (analysis from 3DVar and 3DEnsVar experi-
371 ments) were made using the same version of the static background error covariance matrix.
372 In Section a, it was showed how the static part of the hybrid background error covariance
373 matrix was calculated and how its application within GSI is made. Following the results in
374 Wu et al. (2002), the main features of the static **B** are presented here in a similar manner.

375 In order to make an quantitative and a qualitative description of the amplitudes represented
376 by the computed covariance matrix, Figure 2 depicts the standard deviations of the variance
377 error for the ψ , χ , T , rh , oz and ps . As a reference, Table 4 presents the maximum and

378 minimum values for the amplitudes shown in Figure 2 along with the values for the horizontal
379 and vertical length-scales.

380 The distribution of the variances of ψ and χ are presented in a vertical cross-section, as
381 shown in Figure 2. These distribution shows some of the features of the non-divergent and
382 divergent part of the wind for the BAMv0 model. The ψ variances distribution are sort of
383 symmetric with respect to the Equator line. They are more elongated throughout the vertical
384 column and are concentrated at the model top. Higher values of the streamfunction variances
385 are located over the poles, between sigma levels of 0.6 (~ 614.4 hPa) and 0.2 (~ 204.8 hPa).
386 In the tropical regional, these values are more confined between 0.3 (~ 307.2 hPa) and 0.1
387 (~ 102.4 hPa) sigma levels. The distribution of the variances of the velocity potential also
388 shows a symmetric pattern with respect to the Equator line and are mostly concentrated
389 with higher values over the tropical region, between sigma levels of 0.3 (~ 307.2 hPa) and 0.1
390 (~ 102.4 hPa).

391 Temperature variances are distribute from the Equator line towards the pole, in both
392 hemispheres. At 30°N large values extend onwards with a maximum peak between 0.9 (~ 921.6
393 hPa) and 0.8 (~ 819.2 hPa) sigma layers. At 60°S , a symmetric pattern is represented, as can
394 be seen in Figure 4c in Wu et al. (2002). Variances of the humidity are distributed with a
395 peak in approximately 60°S between 0.7 (~ 716.8 hPa) and 0.6 (~ 614.4 hPa) sigma layers.
396 Layers above 0.6 sigma, variances diminish with height until layer 0.2 (~ 204.8 hPa), whereas
397 the opposite pattern seems to happen around latitude 60°N . Ozone distribution of variances
398 are concentrated in upper layers, 0.5 (~ 512 hPa) upwards, with a peak in the tropical region.
399 In the current setup, GSI does not analyses ozone observations. In the BAMv0 model ozone
400 is transported throughout the grid during the forecast. The surface pressure variances are
401 distributed with two local maximum values in the midlatitudes (higher values are located
402 below 60°S) and minimum values in the tropical region. It is worth to note that the surface
403 pressure variance distribution are fairly different from what can be found in Figure 4e from
404 Wu et al. (2002).

405 At higher resolution (e.g., BAMv0 TQ0299L064), the features described here are less
406 smooth and more detailed (figured not shown), considering a similar model setup but suited
407 for a better resolution in both horizontal and vertical. As a matter of fact, most of the
408 characteristics in the distribution of the error amplitudes may be due to the representation of
409 the BAMv0 topography, including the choices for the model physics and the treatment of the
410 model dynamics and resolution. At a lower resolution (e.g., TQ0062L028), steep topography
411 may be not well represented (peaks in topography may be too smooth), what can lead to
412 discrepancies in the representation of the pressure and winds.

413 A separated study still has to be made in order to address the sensitivity of the background
414 error covariance matrix to the BAMv0 options and the resulting variances. Furthermore, one
415 may consider the fact that such amplitudes can be tuned within GSI by setting up specific
416 parameters for the application of the covariances using recursive filters.

417 Table 4 summarizes the maximum and minimum values for the amplitudes of the standard
418 deviations and horizontal and vertical length-scales, including for the other quantities not
419 shown in Figure 2.

420 *b. Ensemble Innovation Statistics*

421 For an evaluation of the ensemble of analyses produced in experiments using the EnKF
422 and EnSRF, an ensemble innovation assessment is made in order to diagnose deficiencies in
423 the ensemble spread due to inflation and localization adjustment.

424 To assess how the ensemble deals with the system spread in the presence of observational
425 error, innovations are shown as a measure of how good the ensemble spread is due to the
426 observation innovation. In the next figures (Figures 3 and 4), a comparison between the
427 innovation statistics from EnKF and EnSRF is shown for three different regions (NH; TR
428 and SH) and only for the conventional observations. We left the same evaluation regarding
429 unconventional observations for a separated study. In order to make it easier to understand,
430 the y axis shows the ratio between the standard deviation of the prior innovations (i.e.,

431 $y^o - \mathbf{H}x_k^b$, where k is an ensemble member) and the square root of the ensemble total spread
432 (i.e., $\sqrt{\mathbf{S} + \mathbf{R}}$). The ensemble innovation is given by,

$$EI = \frac{y^o - \mathbf{H}x_k^b}{\sqrt{\mathbf{S} + \mathbf{R}}} \quad (4)$$

433 where,

- 434 • y^o : is the observation vector;
- 435 • x_k^b : is the vector of the k -th ensemble member;
- 436 • \mathbf{S} : is the ensemble spread;
- 437 • \mathbf{R} : is the observation error from the observation error covariance matrix.

438 As we are normalizing the standard deviation of the priors innovations by the ensemble
439 total spread, the lower the values, the better the ensemble innovation is. Ideally, the standard
440 deviation of the priors should match the value of the ensemble total spread if the ensemble is
441 well conditioned, i.e., with a good amount of spread and proper localization and inflation
442 values. In our case, as we have tested two different versions of the ensemble Kalman Filter,
443 we kept the same ensemble configurations in all experiments.

444 In Figures 3 and 4, the upper panel shows the statistics for 00UTC and the bottom panel
445 the same statistics for 12UTC. This was made with purpose to reduce the size of the time
446 series and to separate the signal from both times, so this would make easier to analyze the
447 results. Furthermore, Synoptic times of 06 and 18UTC were kept away from the evaluation
448 due to the fact that there is less meteorological observations at these synoptic times. Red
449 lines (dashed and solid) refers to the hybrid 3DEnsVar experiment where 50% of ensemble
450 contribution to the static part of the covariance matrix, whereas the blue lines refers to the
451 same experiment but denoting the case where 75% of ensemble contribution to the static
452 part of the covariances were drawn.

453 Figure 3 shows the ensemble innovation statistics for the hybrid 3DEnsVar experiment
454 using the EnKF algorithm. In general, there is not much difference between the synoptic
455 times of 00 and 12 UTC, but one can say that the amplitudes of the priors (i.e., the ensemble
456 of backgrounds) and posteriors (i.e., the ensemble of posteriors) varies slightly between them.
457 This may be due to the variations of the observations availability at 06 and 18UTC. For
458 some variables, the standard deviation of the priors seemed really far from matching the
459 ensemble total spread. For example, temperature (T - Figures 3d,e,f) at all regions saturate
460 the standard deviation of the priors and them keeps with high values until the end of the
461 simulations. In the region TR, T (Figure 3e) took roughly 20 cycles (i.e., 5 days) to stabilize
462 the ensemble innovation statistics. A similar situation happens in the region SH for T (Figure
463 3f) and also for the region TR for variable q (Figure 3h). It shows that the standard deviation
464 of the priors is greater than the ensemble total spread and is increasing in time. In other
465 words, it mean that the background error seems to dominate the ensemble spread and it
466 shows a lack of proper inflation in the ensemble for these variable in these regions.

467 Surface pressure (ps) in the regions NH (Figure 3m) shows that the standard deviation of
468 the priors tends to increase in time, but it is compensated by the ensemble spread, indicating
469 that the ensemble tries to adjust the observation information properly in the presence of
470 ensemble spread. A different situation happens in the TR and SH where the standard
471 deviation of the priors for ps (Figures 3n,o) is also greater than the ensemble spread, but
472 rather is also negative. This negative signal indicates that ensemble innovation (i.e., $y^o - \mathbf{H}x_k^e$)
473 is negative with the background trying to correct the observations.

474 Differences between the relative amount of ensemble contribution seems to be more or
475 less sensitive to some variables. For ps at TR and SH (Figures 3n,o) it is clear that the 50%
476 of EnKF contribution to the static part of the covariances does a good job if compared to
477 the 75% of ensemble contribution from EnSRF, regardless the synoptic time (00 or 12UTC).
478 This may be due to the observation perturbation made by the EnKF algorithm when the
479 ensemble background plays a major role in the determination of the standard deviation of

480 the innovations.

481 Figure 4 shows the same results of Figure 3, but for the hybrid 3DEnsVar experiments
482 with the EnSRF algorithm. In general, results for the hybrid using the EnSRF indicates that
483 the differences between the priors and posteriors are greater than the corresponding ones
484 using the EnKF. Figures 4d,e,f with the ensemble innovation statistics for T at NH, TR and
485 SH, regardless the relative amount of ensemble contribution, shows that these differences are
486 more noticeable: at the NH, T is closer to the zero line with the EnSRF than with the EnKF.

487 The standard deviation for the horizontal wind (uv) at 00UTC (Figure 4a), was very close
488 to the ensemble total spread, indicating that the ensemble fit to the observation was good
489 and the amount of ensemble spread was reasonable. At the TR region (Figure 4b), however,
490 the opposite was found. In the SH region (Figure 4c), a similar situation to HN happens and
491 the amplitude of the ensemble innovation signal increases indicating, possibly, that lack of
492 ensemble inflation due to reduced number of conventional observation in the region.

493 *c. Forecast Skill*

494 The forecast skill up to 5 days, was assessed in terms of the Anomaly Correlation (AC).
495 As CPTEC developments are focused in the Tropics and the the South America regions,
496 results are shown for these region along with for the whole Globe and North and South
497 Hemispheres. To assess how different the experiments are between each other, a t-Student
498 test was also made with a 95% of confidence level. For each plot of the AC, the t-Student
499 test is presented and the lower boxes shows whether or not the curves are different at the
500 defined confidence level. When the curves crosses their respective boxes, it means that the
501 differences between the experiment (e.g., EnKF50) and the experiment of reference (i.e.,
502 REF, the BAMv0 model integrated with the operational NCEP analysis) are of no difference,
503 indicating that the test for the null hypothesis fails. In this case, the null hypothesis is that
504 the mean of the hybrid experiment verified is statistically indistinguishable from the mean of
505 the experiment of reference.

506 Figure 5 shows the 5-day forecasts AC with a t-Student test for regions NH, TR and SH.
507 Figure 6 shows the 5-day forecasts AC with the same t-Student test, but for the regions GL
508 and SA. Variables evaluated are the surface pressure (*psnm*), specific humidity at 925 hPa
509 (*q925*), air temperature at 850 hPa (*T850*), zonal wind component at 250 hPa (*u250*) and
510 geopotential height at 500 hPa (*z500*). Experiments summarized in Table 1 were run with
511 the same configuration (i.e., using the same code and options chosen through namelists).
512 Experiments EnKF75 and EnSRF50 indicate the best analysis performance for all regions
513 and variables evaluated. Surface pressure (*psnm*) at NH has practically no difference using
514 either EnKF75 or EnSRF50 for the first 24 hours of forecasts. This is an indication that
515 for the 3DEnsVar surface pressure analysis, the EnSRF algorithm with 50% of ensemble
516 contribution, a smaller contribution from the ensemble was utilized. This can be due to the fact
517 that observations used within EnSRF were not perturbed, which in turn could be beneficial
518 to the hybrid analysis tested in this very specific system.

519 In the TR region - a region specially difficult to predict for, for the forecast of *psnm*,
520 *q925* and *T850* (Figures 5b,e,h), both EnKF75 and EnKF75 have shown a good performance
521 with respect experiments REF and 3DVar. The improvements with respect to REF for
522 *psnm* (Figure 5b), are of almost 24 hours (considering 80% of the AC). Figure 5e, shows
523 improvements of the experiment EnSRF75 with respect to experiment 3DVar, in the order of
524 almost 4 days in advance (considering 85% of the AC). A similar improvement was found for
525 the forecast of the *T850* (Figure 5h), with respect to experiment to experiment 3DVar were
526 experiment EnSRF75 still holds 80% of AC for 72 hours forecast, whereas experiment 3DVar
527 limits its forecast skill for this variable up to 36 hours. It is worth to note in Figures 5b,h,n
528 the spin up time that the BAMv0 model took to initialize with the GFS-NCEP analysis.
529 However, this effect was not noticed for every evaluated variable. Figure 5g, for example,
530 indicates that there is no practical difference between the experiments whose AC scores are
531 very similar. The t-Student test, however, reveals that the 3DVar experiment did not fail the
532 null hypothesis test for the 120 hours forecasts, while the remaining of the experiments holds

533 the test up to 36 hours (EnKF50 and EnSRF50) and 72 hours (EnKF75 and EnSRF75).
534 On the other hand, however, experiments with 75% of ensemble contribution for the static
535 covariance, were statistically different from the REF up to 72 hours.

536 Figure 6 shows the same evaluation as Figure 5, but for regions GL and SA. The SA
537 region is of great interest for CPTEC because it is Brazil's target forecast region. The same
538 response from experiments EnKF75 and EnSRF50 were found, indicating that for the SA
539 region, the ensemble covariances can also play an important role in the determination of the
540 forecasts skill. For the GL region, AC scores are similar to what we have found on NH region.
541 Besides that, the AC for $q925$ at GL region (Figure 6c), shows that the forecasts drawn from
542 the EnSRF75 experiment (together with EnKF75) were good for 120 hours. In this case,
543 however, the t-Student test revealed that experiment EnKF75 fails the null hypothesis test
544 at 72 hours forecast, when the REF analysis made the BAMv0 model to produce forecasts
545 with better correlation with the climatology. Remaining variables (i.e., $psnm$, $T850$, $u250$
546 and $z500$ Figures 6a,e,g,i, respectively) showed similar AC scores in comparison with the
547 NH region. For the SA region, good results were found $psnm$ (Figure 6b), for up to 84 hours
548 forecasts; $q925$ (Figure 6d), with AC greater than 85% for 120 hours and $T850$ (Figure 6f)
549 with AC of 80% from experiment EnSRF75 for up 72 hours forecasts and from experiment
550 EnKF75 for up to 60 hours forecasts. Zonal wind at 250 hPa $u250$ and, specially, $z500$, did
551 not drawn good AC scores. The $z500$ from experiments EnKF75 and EnSRF75 did not
552 performed as for others variables. From 84 hours forecast ahead, they were the worse and
553 experiments EnKF50, EnSRF50 and even experiment 3DVar were better.

554 A comparison between the analysis schemes used in the experiments, shows that the
555 humidity analysis drawn from the EnKF75 and EnSRF50 experiments, were improved from
556 the pure 3DVar analysis. Figures 5c, 5f, 5i and 6c and 6f shows that the humidity forecasts
557 were worse even with the data assimilation cycle when compared with the use of the REF
558 analysis. For these hybrid 3DEnsVar experiments, background error covariance matrix is the
559 same static covariance matrix used within the 3DVar experiment. This is an indication that

560 the ensemble covariance plays an important role defining how the analysis increments are
561 applied.

562 PRECIPITATION ASSESSMENT

563 Precipitation forecasts drawn from the REF experiment and from the conventional 3DVar
564 and the 3DEnsVar analysis has been also evaluated. Figures 7 shows the observed Global
565 Precipitation Climatology Project v2.2 (GPCPv2.2) 2.5° (Adler et al. 2003) global monthly
566 mean precipitation and Figure 8 the forecasted precipitation - represented as a monthly
567 mean, drawn from the REF, 3DVar and the four 3DEnsVar hybrid analyses, as summarized
568 in Table 1. The comparison between the 24 hour monthly mean forecasts and the GPCP 2.5
569 is valid for January 2013 (12UTC) and the gridded data has been interpolated to a 2.5° grid
570 resolution to match GPCPv2.2. GPCP 24 hour monthly mean precipitation for January 2013
571 depicts the main features of the tropical precipitation (from 8 mm above) and the subtropical
572 precipitation (until 8 mm), with most of the convective precipitation being distributed over
573 the tropical region.

574 Figure 9 shows a comparison between the forecasted precipitation from the various
575 experiments with the hybrid and non-hybrid background error covariance matrix. Apart
576 from experiment REF, all experiments were ran in a data assimilation cycle using the same
577 observational data set (as shown in Table 2), at the same resolution and with the same
578 model setup (as shown in Table 3). The precipitation fields being compared in Figure 7
579 represents the monthly mean at 12UTC beginning at 1st January 2013, until 31rd January
580 2013. All experiments shows the spatial features of the large scale precipitation and the
581 convective precipitation as well. Figure 7 is used as a reference and the evaluation is made
582 by comparing the precipitation intensity, distribution and the spatial average of the monthly
583 mean, annotated below each picture.

584 The precipitation forecasts from the experiments, shows a reasonable spatial distribution
585 in comparison with GPCPv2.2 monthly mean precipitation (Figure 7). It is worth to note

586 that the model configuration for large scale condensation, Cu convection and horizontal
587 and vertical diffusion were kept the same, so the differences are supposed to be due to the
588 initial condition only. Main differences between precipitation forecasts from experiments are
589 mostly related to the intensity of the precipitation, once the spatial distribution between
590 them are fairly similar. Figure 8a, shows the resulting monthly mean 24 hour precipitation
591 from the BAMv0 model initialized with the REF analysis. In comparison with any other
592 experiment (including the GPCPv2.2 reference), the BAMv0 model tended to produce more
593 convective precipitation in the Tropical region. The spatial average of the time mean amounts
594 to 2.9718 mm/month, yielding the highest value among the experiments, even higher than
595 GPCPv2.2 (2.7197 mm/month). All experiments with data assimilation cycle, on the other
596 hand, have produced precipitation with less concentrated distribution and a more reasonable
597 spatial average. The spatial average in experiment EnSRF50 (Figure 8e) is the one which
598 more approximates GPCPv2.2, although the rest of the data assimilation experiments has
599 concentrated the convective precipitation in a more conformable way.

600 The exaggerated amount of precipitation produced by experiment REF must be due
601 to fact that this experiment was made by just running the BAMv0 model with the GFS-
602 NCEP analysis, although a topography smoothing and a initialization with normal modes
603 has been made for every analysis used in this specific experiment. The choices for the
604 Cu parametrization and the diffusion parameters may also play an important role in the
605 precipitation production. The experiments with the data assimilation cycle, with the exception
606 of experiment EnSRF75 (Figure 8f) - which is the experiment that yielded the lower monthly
607 mean, seems to be well conditioned with the option chosen for the forecasts.

608 In order to better understand and to identify the regions where the experiments have
609 accumulated more or less the precipitation, Figure 9 shows the time series of the spatial
610 averages for each experiment, including the GPCP. In this figure, the GPCPv1.2 1DD (1
611 Degree Daily - Huffman et al. (2001)) was used and the observed data has been interpolated
612 to the analysis grid (with 1.875°). Accompanying Figure 9, Tables 5 and 6 shows the spatial

averages (μ) and the standard deviations (σ) for each experiment (including the interpolated GPCP1DD) at the regions evaluated (GL, NH, TR, SH and SA).

In contrast to what we have found by analyzing the spatial distribution of the precipitation forecasts in Figure (8), the time series for the spatial averages showed in Figure 9, shows us that - with the exception of region NH (Figure 8b), all experiments have some difficulties to reproduce the observed precipitation in 24 hour forecasts. Deficiencies can be noted with respect to the amplitude (maxima and minima) and to the average. In some specific cases, forecasted precipitation nearly matched observed values - e.g., experiment EnSRF75 in TR region (Figure 9c) which matched the GPCPv1.2 at during some periods of the month. This results is with accordance to what we found when analyzing the forecast skill (Figures 5 and 6), showing that among the hybrid 3DEnsVar experiments, the 75% of the EnSRF ensemble contribution to the static part of the background error covariance was beneficial in most cases. On the other hand, experiment REF have exaggerated the precipitation over time mainly for the GL and TR regions (Figures 9a and 9c). This is an expected result, since this experiment have used an independent analysis and the model was not configured to give the best result from it. For the TR and SA regions (Figures 9c,e), precipitation forecasts were mostly above the observed values indicated by GPCPv1.2. In these regions, the convective precipitation plays an important role and models tend to exaggerate the precipitation there. If we consider the SH region (Figure 9d) - which contains more oceanic parts (in opposition to the NH region), models tends to produce more large-scale precipitation whereas the convective precipitation tends to be less important in the results. Consequently, the precipitation forecasts tends to be below the observed values.

All experiments with the hybrid background error covariance matrix showed reasonable results if compared to experiments 3DVar and REF. It appears, however, that the precipitation - specially the convective precipitation, to be not so sensitive to changes in the covariance matrix, although the experiments with greater ensemble contributions (e.g., EnKF75 and EnSRF75) do perform better than the rest of the experiments made.

640 5. Summary and discussion

641 In this work we have tested an implementation of the hybrid ensemble-variational
642 (3DEnsVar) data assimilation technique with the BAMv0 model. For this implementa-
643 tion, we carried out the calculation of a static background error covariance matrix used in the
644 linear combination with the covariances from the ensemble part. To test the implementation,
645 experiments were made in a TQ0062L028 model and analysis resolution using an ensemble of
646 analysis of 40 members, testing both the EnKF and EnSRF algorithms. Results shows an
647 general improvement of the winds, temperature, moisture and surface pressure, mainly over
648 South America and the Tropical region, which is a target region for CPTEC operations. The
649 moist part of the BAMv0 model (e.g., Specific Moisture, Precipitable Water and Precipitation)
650 resulted in a slight difference from the control run (the BAMv0 model initialized with the
651 GFS-NCEP analysis) and the from the BAMv0 model simulation initialized with the pure
652 3DVar analysis.

653 In this work it is clear that the ensemble contributions to the static part of the covariance
654 matrix plays an important role in determining the skill of the analysis, and consequently the
655 skill of the forecasts. In some cases, we found improvements of more than 3 days, which we
656 consider a great step towards a robust analysis to use in daily operations. Furthermore, with
657 this implementation, we could not only show that our implementation is in the right way, but
658 also that we could improve the pure 3DVar analysis (despite some practical issues that have
659 to be properly addressed). Our results corroborates and contribute to the understanding of
660 the application of hybrid background error covariances matrices.

661 Further improvements to this work shall include the treatment of the caveats pointed out
662 in Section c. The realization of the 3DEnsVar system at CPTEC operations will also depends
663 on the CPTEC capacity to keep the system at a higher resolution without compromise
664 the system performance and to afford the computational cost involved. A study regarding
665 the dual resolution feature for the hybrid system must be considered as the resolution of
666 the ensemble limits the system performance, and the use of the re-centering and ensemble

⁶⁶⁷ inflation.

⁶⁶⁸ *Acknowledgments.*

⁶⁶⁹ The authors are thankful to Ricardo Todling and Stephen Cohn from GMAO/NASA for
⁶⁷⁰ the guidance and multiple discussions during the implementation of the hybrid system at
⁶⁷¹ CPTEC and Daryl Kleist from NOAA/NCEP. The authors are also thankful to the CPTEC
⁶⁷² staff and to the anonymous reviewers for the corrections and the suggestions that improved
⁶⁷³ the manuscript. To CAPES for the financial support in the PDSE program (grant number
⁶⁷⁴ BEX 99999.008036/2014-04).

REFERENCES

- 677 Adler, R. F., et al., 2003: The Version–2 Global Precipitation Climatology Project (GPCP)
 678 Monthly Precipitation Analysis (1979–Present). *Journal of Hydrometeorology*, **4** (6),
 679 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2, URL [http://dx.doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).
- 680 Berre, L., M. Monteiro, and C. Pires, 2013: An Impact Study of Updating Background Error
 681 Covariances in the ALADIN-France Data Assimilation System. *Journal of Geophysical
 682 Research: Atmospheres*, **118** (19), 11,075–11,086, doi:10.1002/jgrd.50847, URL <http://dx.doi.org/10.1002/jgrd.50847>.
- 683 Cavalcanti, I. F. d. A., et al., 2002: Global Climatological Features in a Simulation using
 684 the CPTEC-COLA AGCM. *Journal of Climate*, **15** (21), 2965–2988, URL <http://urlib.net/sid.inpe.br/iris@1915/2005/05.05.11.19>.
- 685 Chou, M., 1999: *A Solar Radiation Parameterization for Atmospheric Studies*. NASA technical
 686 memorandum, URL https://books.google.com.br/books?id=_gYbvgAACAAJ.
- 687 Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational Implementation of a
 688 Hybrid Ensemble/4D–Var Global Data Assimilation System at the Met Office. *Quarterly
 689 Journal of the Royal Meteorological Society*, **139** (675), 1445–1461, doi:10.1002/qj.2054,
 690 URL <http://dx.doi.org/10.1002/qj.2054>.
- 691 Corazza, M., et al., 2003: Use of the Breeding Technique to Estimate the Structure of the
 692 Analysis “errors of the day”. *Nonlinear Processes in Geophysics*, **10** (3), 233–243, doi:
 693 10.5194/npg-10-233-2003, URL <http://www.nonlin-processes-geophys.net/10/233/2003/>.

- 699 Cunningham, C., B. J. P., and M. Ferreira, 2015: Assessing Improved CPTEC Probabilistic
700 Forecasts on Medium-Range Timescale. *Meteorological Applications*, **22** (3), 378–384,
701 doi:10.1002/met.1464, URL <http://dx.doi.org/10.1002/met.1464>.
- 702 Desroziers, G., 1997: A Coordinate Change for Data Assimilation in Spherical
703 Geometry of Frontal Structures. *Monthly Weather Review*, **125** (11), 3030–3038,
704 doi:10.1175/1520-0493(1997)125<3030:ACCFDA>2.0.CO;2, URL <http://dx.doi.org/>
705 10.1175/1520-0493(1997)125<3030:ACCFDA>2.0.CO;2, [http://dx.doi.org/10.1175/1520-0493\(1997\)125<3030:ACCFDA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1997)125<3030:ACCFDA>2.0.CO;2).
- 706
- 707 Evensen, G., 1994: Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model
708 using Monte Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research: Oceans*, **99** (C5), 10 143–10 162, doi:10.1029/94JC00572, URL <http://dx.doi.org/10.1029/94JC00572>.
- 709
- 710
- 711 Figueroa, S. N., et al., 2016: The Brazilian Global Atmospheric Model (BAM): Performance
712 for Tropical Rainfall Forecasting and Sensitivity to Convective Scheme and Horizontal Res-
713 olution. *Weather and Forecasting*, **31** (5), 1547–1572, doi:10.1175/WAF-D-16-0062.1,
714 URL <http://dx.doi.org/10.1175/WAF-D-16-0062.1>, <http://dx.doi.org/10.1175/WAF-D-16-0062.1>.
- 715
- 716 Fisher, M., 2003: Background Error Covariance Modelling. *Seminar on Recent Developments*
717 in Data Assimilation for Atmosphere and Ocean, 8-12 September 2003, ECMWF, Shinfield
718 Park, Reading, ECMWF, 45–64.
- 719
- 720 Grell, G. A., 1993: Prognostic Evaluation of Assumptions Used by Cumulus Parameteriza-
721 tions. *Monthly Weather Review*, **121** (3), 764–787, doi:10.1175/1520-0493(1993)121<0764:
722 PEOAUB>2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0493\(1993\)121<0764:PEOAUB>](http://dx.doi.org/10.1175/1520-0493(1993)121<0764:PEOAUB>)
723 2.0.CO;2,

- 724 Hamill, T. M. and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter–3D Variational Anal-
725 ysis Scheme. *Monthly Weather Review*, **128** (8), 2905–2919, doi:10.1175/1520-0493(2000)
726 128<2905:AHEKFV>2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0493\(2000\)128<2905:AHEKFV>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2).
- 727
- 728 Harshvardhan, R. Davies, D. A. Randall, and T. G. Corsetti, 1987: A Fast Radiation
729 Parameterization for Atmospheric Circulation Models. *Journal of Geophysical Research:*
730 *Atmospheres*, **92** (D1), 1009–1016, doi:10.1029/JD092iD01p01009, URL <http://dx.doi.org/10.1029/JD092iD01p01009>.
- 731
- 732 Hayden, C. M. and R. J. Purser, 1995: Recursive Filter Objective Analysis of Meteo-
733 rological Fields: Applications to NESDIS Operational Processing. *Journal of Applied*
734 *Meteorology*, **34** (1), 3–15, doi:10.1175/1520-0450-34.1.3, URL <http://dx.doi.org/10.1175/1520-0450-34.1.3>.
- 735
- 736 Holtslag, A. A. M. and B. A. Boville, 1993: Local Versus Nonlocal Boundary-
737 Layer Diffusion in a Global Climate Model. *Journal of Climate*, **6** (10), 1825–1842,
738 doi:10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0442\(1993\)006<1825:LVNBLD>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2).
- 739
- 740
- 741 Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce,
742 B. McGavock, and J. Susskind, 2001: Global Precipitation at One-Degree Daily Res-
743 olution from Multisatellite Observations. *Journal of Hydrometeorology*, **2** (1), 36–50,
744 doi:10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2, URL [http://dx.doi.org/10.1175/1525-7541\(2001\)002<0036:GPAODD>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2).
- 745
- 746
- 747 Lorenc, A. C., 1986: Analysis Methods for Numerical Weather Prediction. *Quarterly Journal*

- 748 *of the Royal Meteorological Society*, **112** (474), 1177–1194, doi:10.1002/qj.49711247414,
749 URL <http://dx.doi.org/10.1002/qj.49711247414>.
- 750 Lorenc, A. C., 2003: Modelling of Error Covariances by 4D–Var Data Assimilation. *Quarterly
751 Journal of the Royal Meteorological Society*, **129** (595), 3167–3182, doi:10.1256/qj.02.131,
752 URL <http://dx.doi.org/10.1256/qj.02.131>.
- 753 Mellor, G. L. and T. Yamada, 1974: A Hierarchy of Turbulence Closure Models for
754 Planetary Boundary Layers. *Journal of the Atmospheric Sciences*, **31** (7), 1791–1806,
755 doi:10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0469\(1974\)031<1791:AHOTCM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2), [http://dx.doi.org/10.1175/1520-0469\(1974\)031<1791:AHOTCM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2).
- 756 757
- 758 Mendonça, A. M. and J. P. Bonatti, 2009: Experiments with EOF–Based Perturbation
759 Methods and Their Impact on the CPTEC/INPE Ensemble Prediction System. *Monthly
760 Weather Review*, **137** (4), 1438–1459, doi:10.1175/2008MWR2581.1, URL <http://dx.doi.org/10.1175/2008MWR2581.1>, <http://dx.doi.org/10.1175/2008MWR2581.1>.
- 761
- 762 Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center’s Spectral
763 Statistical-Interpolation Analysis System. *Monthly Weather Review*, **120** (8), 1747–1763,
764 doi:10.1175/1520-0493(1992)120<1747:TNCMSS>2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0493\(1992\)120<1747:TNCMSS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1992)120<1747:TNCMSS>2.0.CO;2).
- 765
- 766 Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003a: Numerical As-
767 pects of the Application of Recursive Filters to Variational Statistical Analysis. Part
768 I: Spatially Homogeneous and Isotropic Gaussian Covariances. *Monthly Weather Review*,
769 **131** (8), 1524–1535, doi:10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2, URL
770 [http://dx.doi.org/10.1175//1520-0493\(2003\)131<1524:NAOTAO>2.0.CO;2](http://dx.doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2), [http://dx.doi.org/10.1175//1520-0493\(2003\)131<1524:NAOTAO>2.0.CO;2](http://dx.doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2).
- 771
- 772 Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003b: Numerical Aspects of

- 773 the Application of Recursive Filters to Variational Statistical Analysis. Part II: Spatially
774 Inhomogeneous and Anisotropic General Covariances. *Monthly Weather Review*, **131** (8),
775 1536–1548, doi:10.1175//2543.1, URL <http://dx.doi.org/10.1175//2543.1>.
- 776 Rabier, F., 2005: Overview of Global Data Assimilation Developments in Numerical Weather-
777 Prediction Centres. *Quarterly Journal of the Royal Meteorological Society*, **131** (613),
778 3215–3233, doi:10.1256/qj.05.129, URL <http://dx.doi.org/10.1256/qj.05.129>.
- 779 Riishøjgaard, L. P., 1998: A Direct Way of Specifying Flow-Dependent Background Error
780 Correlations for Meteorological Analysis Systems. *Tellus A*, **50** (1), 42–57, doi:10.1034/j.
781 1600-0870.1998.00004.x, URL <http://dx.doi.org/10.1034/j.1600-0870.1998.00004>.
- 782 X.
- 783 Sellers, P. J., et al., 1996: A Revised Land Surface Parameterization (SiB2) for At-
784 mospheric GCMs. Part I: Model Formulation. *Journal of Climate*, **9** (4), 676–705,
785 doi:10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2, URL <http://dx.doi.org/>
786 10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2, [http://dx.doi.org/10.1175/1520-0442\(1996\)009<0676:ARLSPF>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2).
- 787
- 788 Thépaut, J.-N. and P. Courtier, 1991: Four-Dimensional Variational Data Assimilation
789 using the Adjoint of a Multilevel Primitive-Equation Model. *Quarterly Journal of the
790 Royal Meteorological Society*, **117** (502), 1225–1254, doi:10.1002/qj.49711750206, URL
791 <http://dx.doi.org/10.1002/qj.49711750206>.
- 792 Tiedtke, M., 1983: The Sensitivity of the Time-Mean Large-Scale Flow to Cumulus Convection
793 in the ECMWF Model. *Workshop on Convection in Large-scale Numerical Models*, 28
794 November to 1 December 1983, ECMWF, Shinfield Park, Reading, ECMWF, 297–316.
- 795 Wang, X., 2010: Incorporating Ensemble Covariance in the Gridpoint Statistical Interpo-
796 lation Variational Minimization: A Mathematical Framework. *Monthly Weather Review*,

- 797 138 (7), 2990–2995, doi:10.1175/2010MWR3245.1, URL <http://dx.doi.org/10.1175/2010MWR3245.1>, <http://dx.doi.org/10.1175/2010MWR3245.1>.
- 798
- 799 Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A Hybrid ETKF–3DVAR
800 Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation
801 Experiment. *Monthly Weather Review*, 136 (12), 5116–5131, doi:10.1175/2008MWR2444.1,
802 URL <http://dx.doi.org/10.1175/2008MWR2444.1>.
- 803 Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008b: A Hybrid ETKF–3DVAR
804 Data Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments.
805 *Monthly Weather Review*, 136 (12), 5132–5147, doi:10.1175/2008MWR2445.1, URL <http://dx.doi.org/10.1175/2008MWR2445.1>.
- 806
- 807 Wang, X., T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007: A Comparison of
808 Hybrid Ensemble Transform Kalman Filter–Optimum Interpolation and Ensemble Square
809 Root Filter Analysis Schemes. *Monthly Weather Review*, 135 (3), 1055–1076, doi:10.
810 1175/MWR3307.1, URL <http://dx.doi.org/10.1175/MWR3307.1>, <http://dx.doi.org/10.1175/MWR3307.1>.
- 811
- 812 Wang, X., T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2009: A Comparison of the
813 Hybrid and EnSRF Analysis Schemes in the Presence of Model Errors due to Unre-
814 solved Scales. *Monthly Weather Review*, 137 (10), 3219–3232, doi:10.1175/2009MWR2923.
815 1, URL <http://dx.doi.org/10.1175/2009MWR2923.1>, <http://dx.doi.org/10.1175/2009MWR2923.1>.
- 816
- 817 Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-Based Ensemble-
818 Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single-
819 Resolution Experiments. *Monthly Weather Review*, 141 (11), 4098–4117, doi:10.1175/
820 MWR-D-12-00141.1, URL <http://dx.doi.org/10.1175/MWR-D-12-00141.1>, <http://dx.doi.org/10.1175/MWR-D-12-00141.1>.
- 821

- 822 Whitaker, J. S. and T. M. Hammill, 2002: Ensemble Data Assimilation With-
823 out Perturbed Observations. *Monthly Weather Review*, **130** (7), 1913–1924, doi:
824 10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2, [http://dx.doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).
- 826 Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-Dimensional Variational Analysis
827 with Spatially Inhomogeneous Covariances. *Monthly Weather Review*, **130** (12), 2905–2916,
828 doi:10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2).
- 830 Zupanski, M., 2005: Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Monthly
831 Weather Review*, **133** (6), 1710–1726, doi:10.1175/MWR2946.1, URL <http://dx.doi.org/10.1175/MWR2946.1>,
832 <http://dx.doi.org/10.1175/MWR2946.1>, <http://dx.doi.org/10.1175/MWR2946.1>.

⁸³⁴ **List of Tables**

⁸³⁵ 1	3DEnsVar experiments and hybrid covariance matrix setup.	36
⁸³⁶ 2	Observational data used in the experiments with a cycled analysis.	37
⁸³⁷ 3	BAMv0 model setup (main options) used within the data assimilation cycles.	38
⁸³⁸ 4	Maximum and minimum values of Amplitudes, Horizontal and Vertical Length-Scales of the Static TQ0062L028 background error covariance matrix.	39
⁸⁴⁰ 5	Area average of 24 hour (12UTC) precipitation (in mm/month) for the Globe (GL) and South America (SA).	40
⁸⁴² 6	Area average of 24 hour (12UTC) precipitation (in mm/month) for the North Hemisphere (NH), Tropics (TR) and South Hemisphere (SH).	41
⁸⁴³		

TABLE 1. 3DEnsVar experiments and hybrid covariance matrix setup.

Experiment	Hybrid Setup	Description
REF	100% Static	B based on GFS-NCEP global forecasts, 1 member (deterministic)
3DVar	100% Static	B based on BAMv0 model, 1 member (deterministic)
EnKF50	50% Static/Ensemble	B based on BAMv0 model, 40 members
EnKF75	75% Ensemble	B based on BAMv0 model, 40 members
EnSRF50	50% Static/Ensemble	B based on BAMv0 model, 40 members
EnSRF75	75% Ensemble	B based on BAMv0 model, 40 members

TABLE 2. Observational data used in the experiments with a cycled analysis.

Mnemonic	Type	Description
airsbufr	Unconventional	AMSU-A/AIRS radiances from AQUA
amsuabufr	Unconventional	AMSU-A 1b radiances (brightness temperature) from NOAA-15, 16, 17, 18, 19 and METOP-A
hirs4bufr	Unconventional	HIRS4 1b radiances from NOAA 18, 19 and METOP-A
mhsbufr	Unconventional	MHS moisture sounders MHS from NOAA 18, 19 and METOP-A
iasibufr	Unconventional	IASI sounders from METOP-A
gpsrobufr	Conventional	GPS radio occultation refractivities
prepbufr	Conventional	$ps, T, q, pw, uv, spd, dw$ and sst observations

TABLE 3. BAMv0 model setup (main options) used within the data assimilation cycles.

Dynamical Setup		Physical Setup	
Resolution	TQ0062L028	LW radiation	Harshvardhan et al. (1987)
Integration timestep	1200 seconds	SW radiation	CLiRAD (Chou 1999)
Wind transport	Eulerian	Cu convection	Grell (1993)
Mass transport	Semi-Lagrangean	Shallow convection	Tiedtke (1983)
Initialization	Diabatic	Top of PBL	Holtslag and Boville (1993)
Restart	Surface	Bottom of PBL	Mellor and Yamada (1974)
Mass conservation	$\ln(p)$	Surface model	SSiB (Sellers et al. 1996)

TABLE 4. Maximum and minimum values of Amplitudes, Horizontal and Vertical Length-Scales of the Static TQ0062L028 background error covariance matrix.

	Amplitudes		Horizontal L-S		Vertical L-S	
	Min	Max	Min	Max	Min	Max
ψ	7.60x10 ⁵	6.83x10 ⁶	4.59x10 ⁵	1.31x10 ⁶	1.00x10 ⁻²	1.48x10 ⁰
χ	0	5.05x10 ⁶	0	1.95x10 ⁶	0	1.60x10 ⁰
t	0	2.43x10 ⁰	0	7.49x10 ⁵	0	1.77x10 ⁰
q	0	6.92x10 ⁻¹	0	7.29x10 ⁻¹	0	6.61x10 ⁵
oz	0	1.76x10 ⁶	0	5.90x10 ⁵	0	5.30x10 ⁰
cw	0	1.95x10 ⁶	0	5.00x10 ⁵	0	4.91x10 ⁰
ps	6.84x10 ⁻²	2.29x10 ⁻¹	3.56x10 ⁵	5.26x10 ⁵	—	—
sst	2.00x10 ⁻²	4.32x10 ⁻¹	5.07x10 ²	8.00x10 ²	—	—

TABLE 5. Area average of 24 hour (12UTC) precipitation (in mm/month) for the Globe (GL) and South America (SA).

	GL		SA	
	μ	σ	μ	σ
GPCP	2.8635	0.1123	3.1321	0.7743
REF	3.0478	0.1036	5.2812	0.7786
3DVar	2.7601	0.1118	5.0152	0.8779
EnKF50	2.6107	0.0999	4.3304	0.7679
EnKF75	2.6620	0.1057	3.8494	0.7618
EnSRF50	2.8103	0.0935	4.6480	0.7768
EnSRF75	2.5745	0.1089	3.8128	0.6702

TABLE 6. Area average of 24 hour (12UTC) precipitation (in mm/month) for the North Hemisphere (NH), Tropics (TR) and South Hemisphere (SH).

	NH		TR		SH	
	μ	σ	μ	σ	μ	σ
GPCP	1.7228	0.2653	4.2820	0.2694	2.6068	0.2053
REF	1.5547	0.2842	5.4655	0.3733	2.0758	0.2685
3DVar	1.4344	0.2438	4.7657	0.3420	2.0499	0.2512
EnKF50	1.4799	0.2576	4.4379	0.2849	1.8863	0.2459
EnKF75	1.4948	0.2744	4.4412	0.2835	2.0280	0.2502
EnSRF50	1.5605	0.2779	4.8385	0.2990	1.9999	0.2546
EnSRF75	1.4566	0.2520	4.2710	0.2209	1.9756	0.2619

844 **List of Figures**

- 845 1 Schematic diagram depicting the 3DEnsVar analysis cycle for the BAMv0 model. 44
846 2 Standard deviations of the variance errors distribution for ψ [$10^6 \text{ m}^2\text{s}^{-1}$], χ
847 [math>10^6 \text{ m}^2\text{s}^{-1}], T [K], q , oz and ps [$\ln(hPa)$] throughout latitudes and vertical
848 sigma levels. 45
- 849 3 Ensemble innovation statistics for the EnKF experiments: upper panel, in-
850 novation statistics valid for 00UTC; bottom panel, innovation statistics valid
851 for 12UTC. Red (blue) color represent hybrid analyses with 50% (75%) of
852 EnKF covariances contribution. Dashed lines represents the priors and solid
853 lines represents the posteriors. Left column represent the North Hemisphere,
854 middle the Tropics and right column, represent the South Hemisphere. 46
- 855 4 Ensemble innovation statistics for the EnSRF experiments: upper panel,
856 innovation statistics valid for 00UTC; bottom panel, innovation statistics valid
857 for 12UTC. Red (blue) color represent hybrid analyses with 50% (75%) of
858 EnSRF covariances contribution. Dashed lines represents the priors and solid
859 lines represents the posteriors. Left column represent the North Hemisphere,
860 middle the Tropics and right column, represent the South Hemisphere. 47
- 861 5 5-day forecast skill for Surface Pressure (first row), Specific Humidity at 850
862 hPa (second row), Absolute Temperature at 850 hPa (fourth row), Horizontal
863 Wind at 250 hPa (fifth row) and Geopotential Height at 500 hPa (sixth row).
864 In the first column are represented the skill for the Northern Hemisphere,
865 second column for the Tropics and the third row for the Southern Hemisphere.
866 All figures are presented with their respective t-Student significance test, where
867 a 95% confidence interval is hold. 48

- 868 6 5-day forecast skill for Surface Pressure (first row), Specific Humidity at 850
869 hPa (second row), Absolute Temperature at 850 hPa (fourth row), Horizontal
870 Wind at 250 hPa (fifth row) and Geopotential Height at 500 hPa (sixth row).
871 In the first column are represented the skill for the whole globe and the second
872 column for the South America. All figures are presented with their respective
873 t-Student significance test, where a 95% confidence interval is hold. 49
- 874 7 Monthly mean precipitation from GPCPv2.2 2.5° valid for January 2013 at
875 12UTC (in mm/month). 50
- 876 8 Monthly average precipitation 24 hour forecast valid for January 2013 at
877 12UTC. Comparison is made between monthly mean precipitation from
878 GPCPv2.2 and all experiments using different analyses. Below each picture,
879 it is annotated the spatial average of the monthly mean of the precipitation
880 forecasts, in mm/month. 51
- 881 9 Spatial averages of the 24 hour forecast precipitation (in mm/day) valid for
882 January 2013 at 12Z. Panel a) precipitation for the whole globe (GL); b) same
883 for the North Hemisphere (NH); c) for Tropics (TR); d) for South Hemisphere
884 (SH) and e) for South America (SA). Precipitation from experiments are
885 compared with the observed precipitation from GPCPv1.2. 52

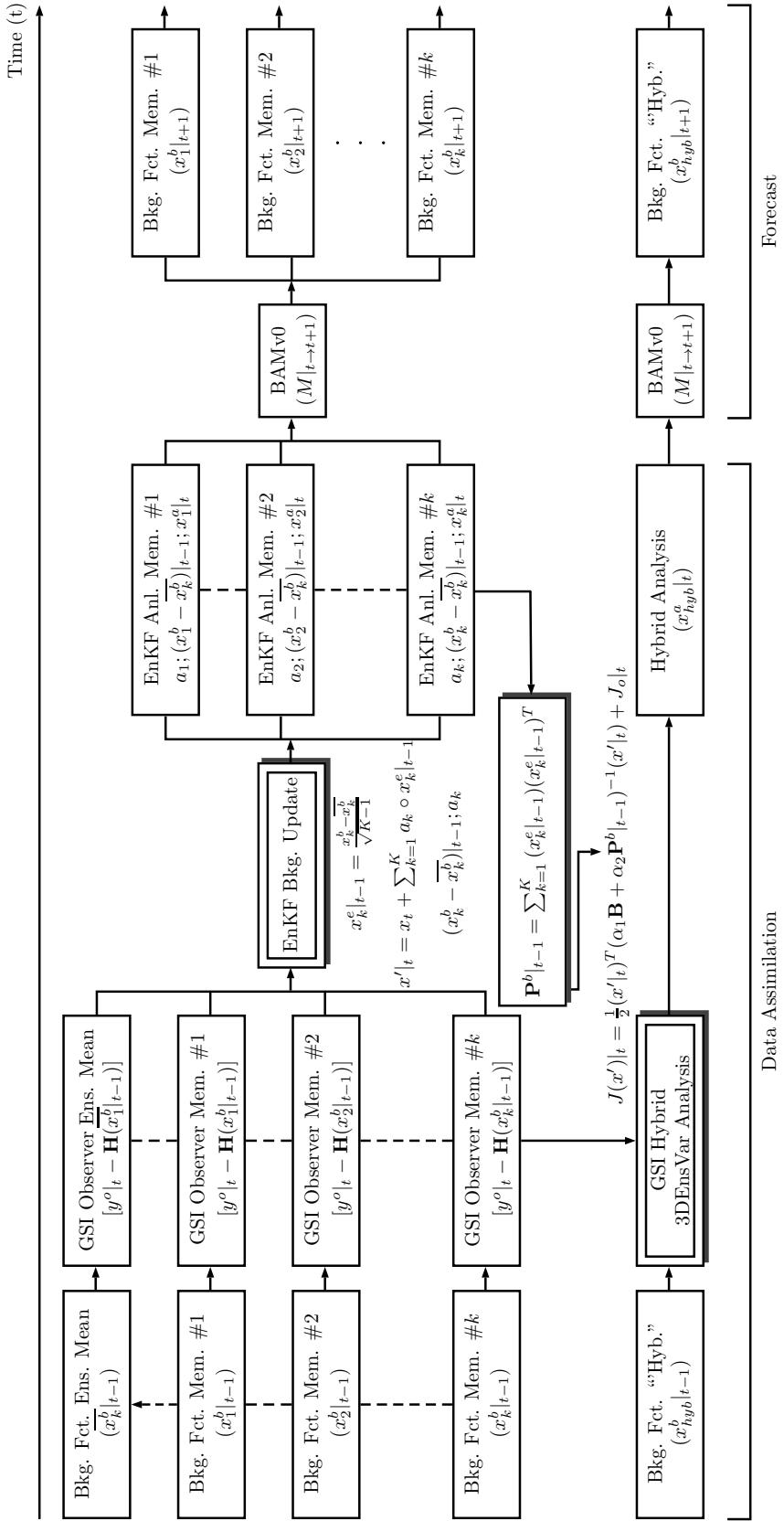


FIG. 1. Schematic diagram depicting the 3DEnSVar analysis cycle for the BAMv0 model.

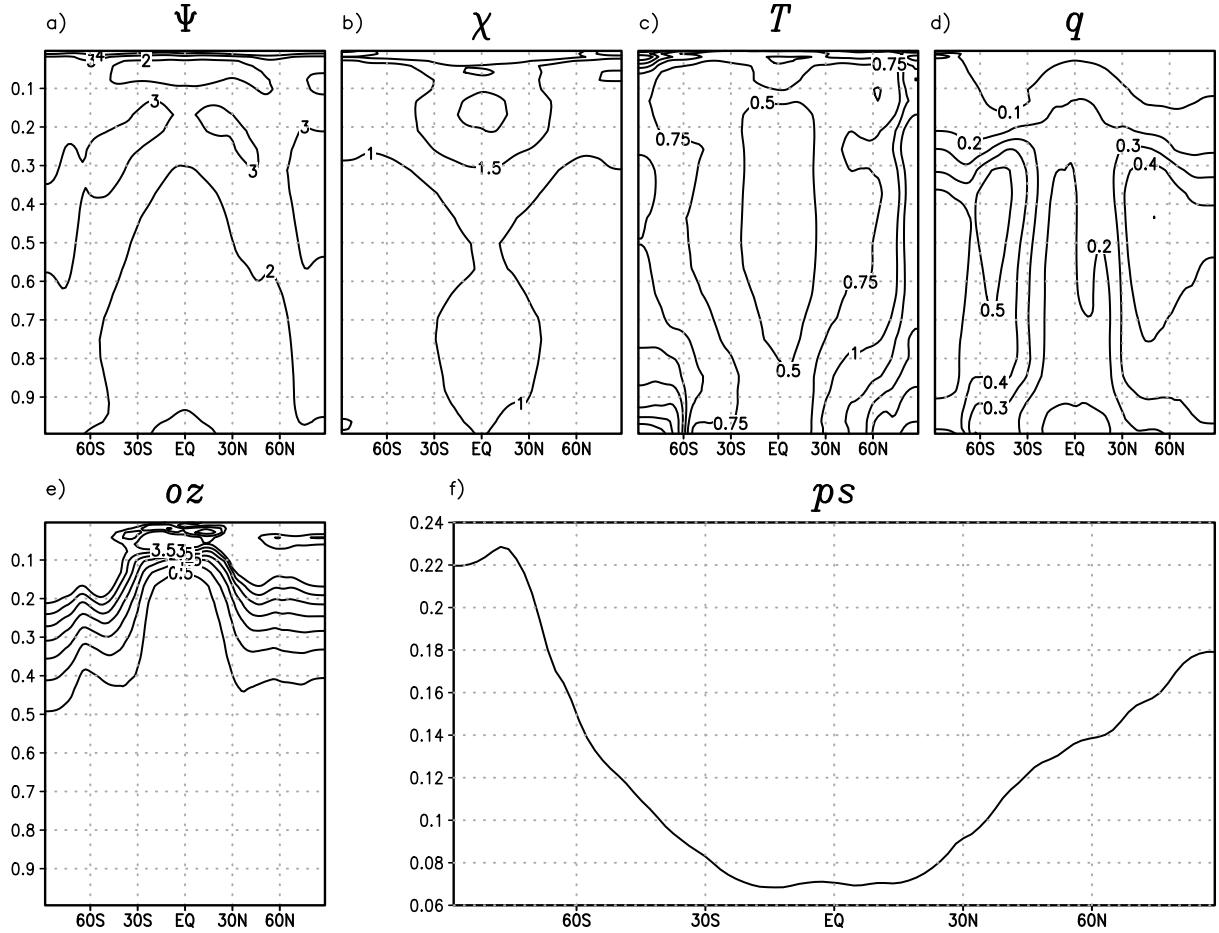


FIG. 2. Standard deviations of the variance errors distribution for ψ [$10^6 \text{ m}^2 \text{s}^{-1}$], χ [$10^6 \text{ m}^2 \text{s}^{-1}$], T [K], q , oz and ps [$\ln(hPa)$] throughout latitudes and vertical sigma levels.

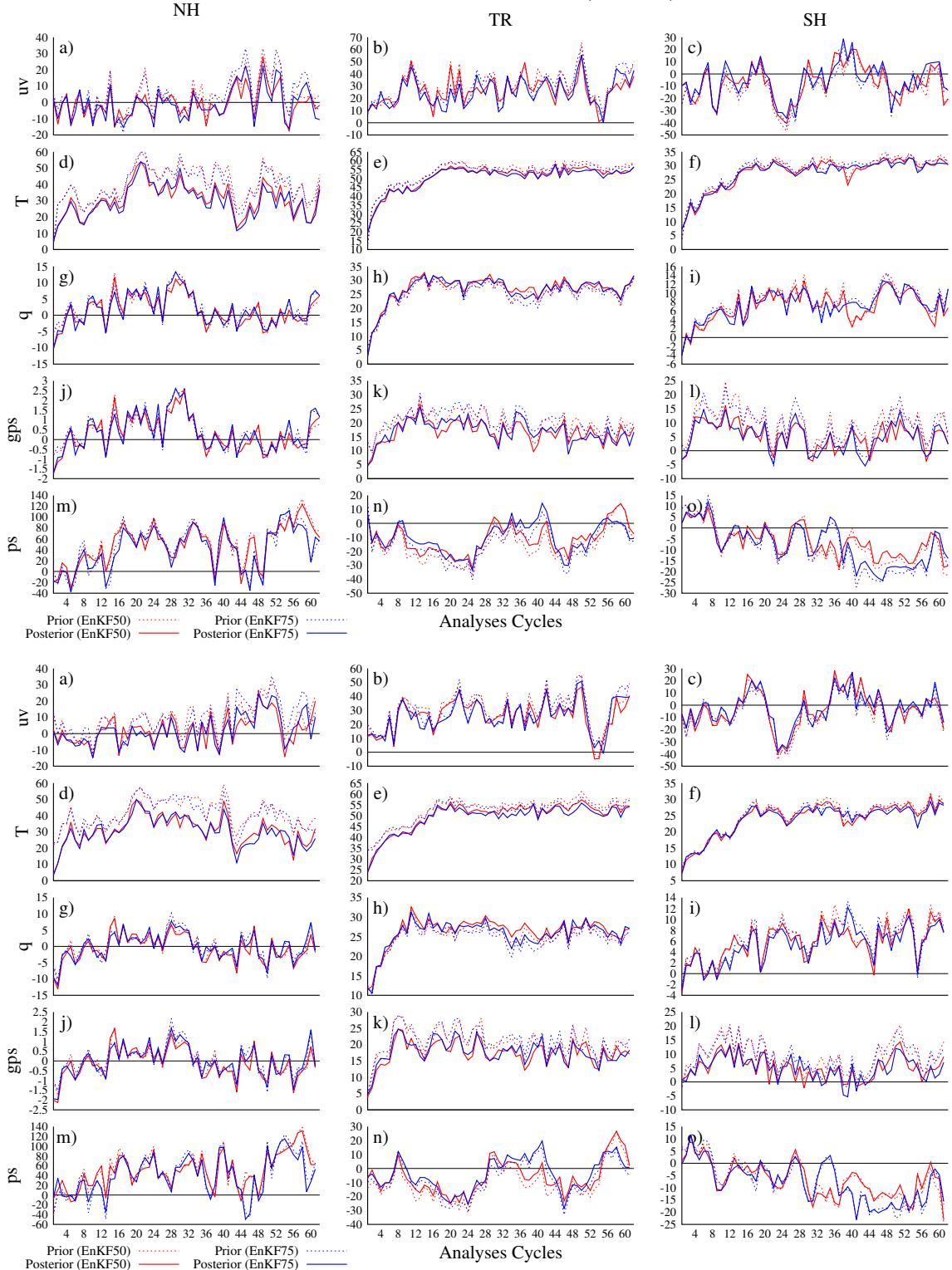


FIG. 3. Ensemble innovation statistics for the EnKF experiments: upper panel, innovation statistics valid for 00UTC; bottom panel, innovation statistics valid for 12UTC. Red (blue) color represent hybrid analyses with 50% (75%) of EnKF covariances contribution. Dashed lines represents the priors and solid lines represents the posteriors. Left column represent the North Hemisphere, middle the Tropics and right column, represent the South Hemisphere.

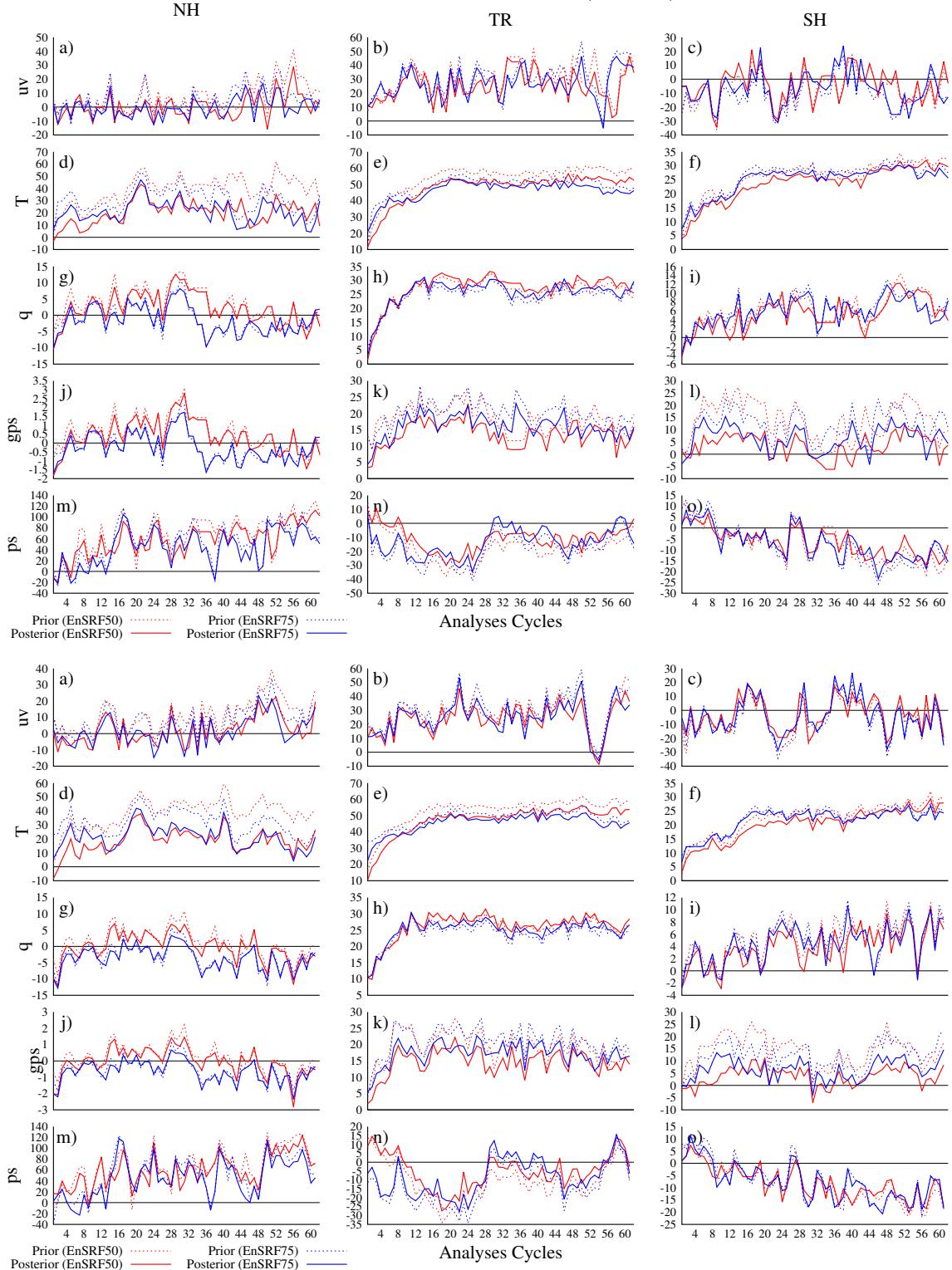


FIG. 4. Ensemble innovation statistics for the EnSRF experiments: upper panel, innovation statistics valid for 00UTC; bottom panel, innovation statistics valid for 12UTC. Red (blue) color represent hybrid analyses with 50% (75%) of EnSRF covariances contribution. Dashed lines represents the priors and solid lines represents the posteriors. Left column represent the North Hemisphere, middle the Tropics and right column, represent the South Hemisphere.

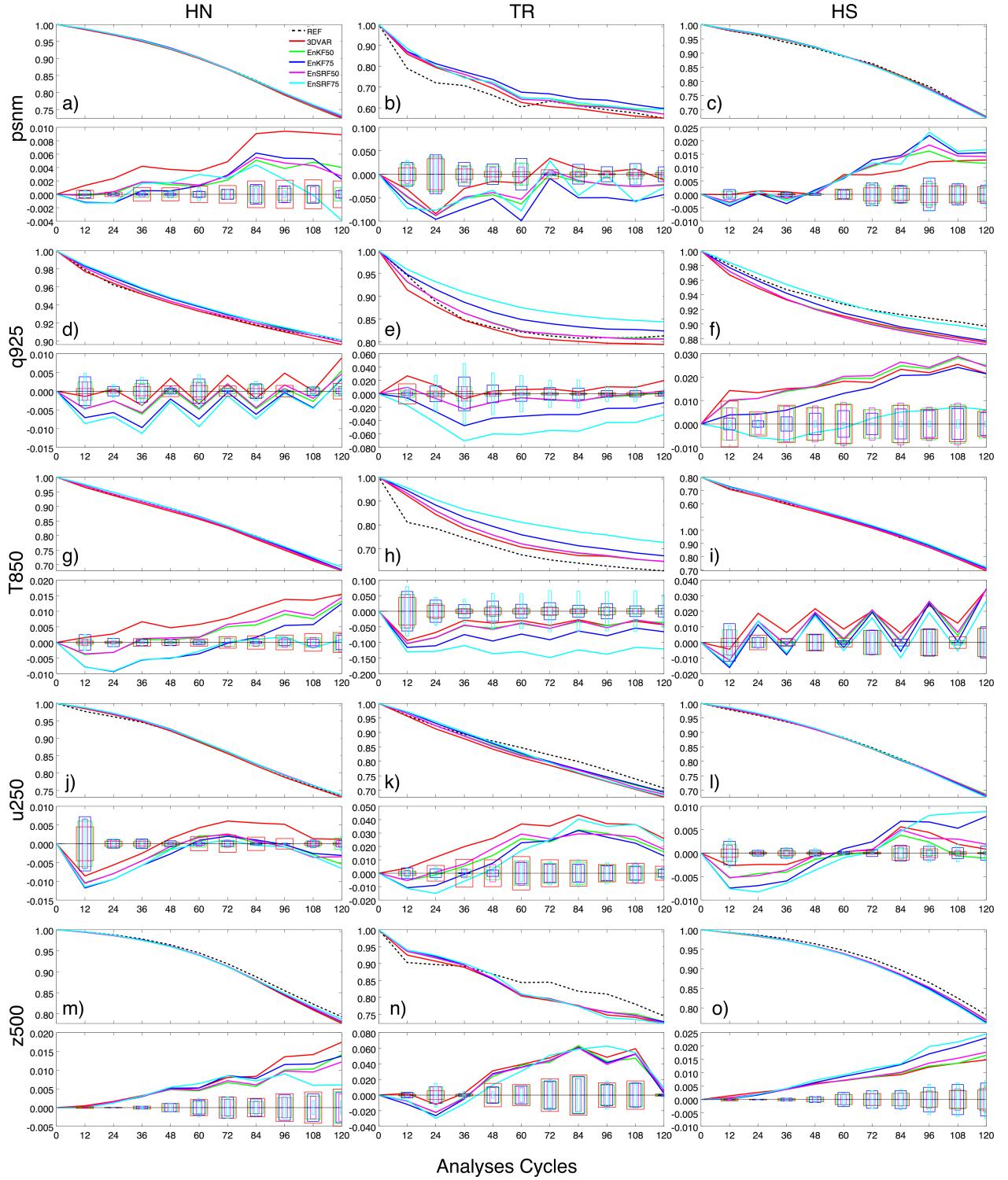


FIG. 5. 5-day forecast skill for Surface Pressure (first row), Specific Humidity at 850 hPa (second row), Absolute Temperature at 850 hPa (fourth row), Horizontal Wind at 250 hPa (fifth row) and Geopotential Height at 500 hPa (sixth row). In the first column are represented the skill for the Northern Hemisphere, second column for the Tropics and the third row for the Southern Hemisphere. All figures are presented with their respective t-Student significance test, where a 95% confidence interval is hold.

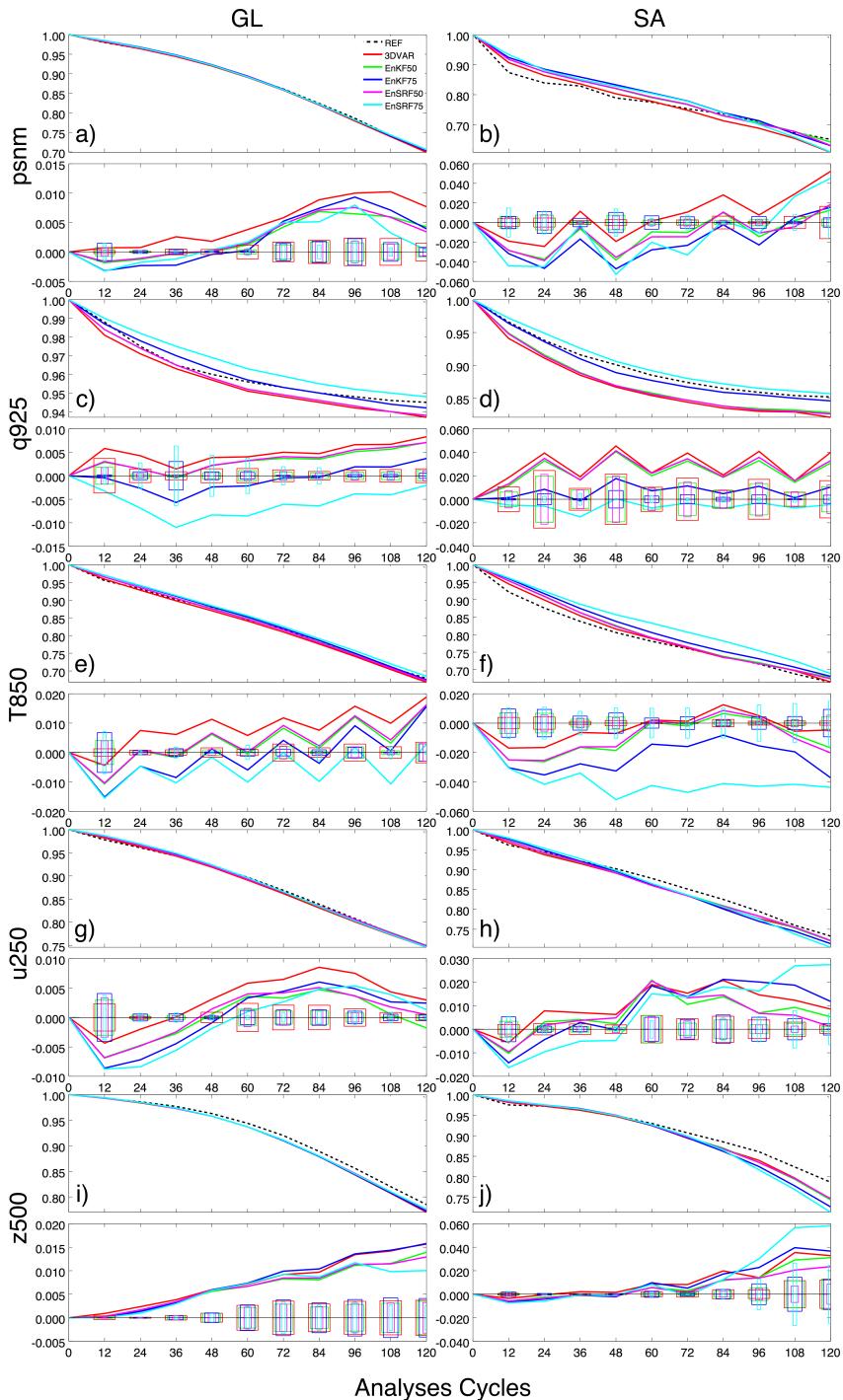


FIG. 6. 5-day forecast skill for Surface Pressure (first row), Specific Humidity at 850 hPa (second row), Absolute Temperature at 850 hPa (fourth row), Horizontal Wind at 250 hPa (fifth row) and Geopotential Height at 500 hPa (sixth row). In the first column are represented the skill for the whole globe and the second column for the South America. All figures are presented with their respective t-Student significance test, where a 95% confidence interval is hold.

GPCP 2.5 24h Precip. Ave. (January 2013)

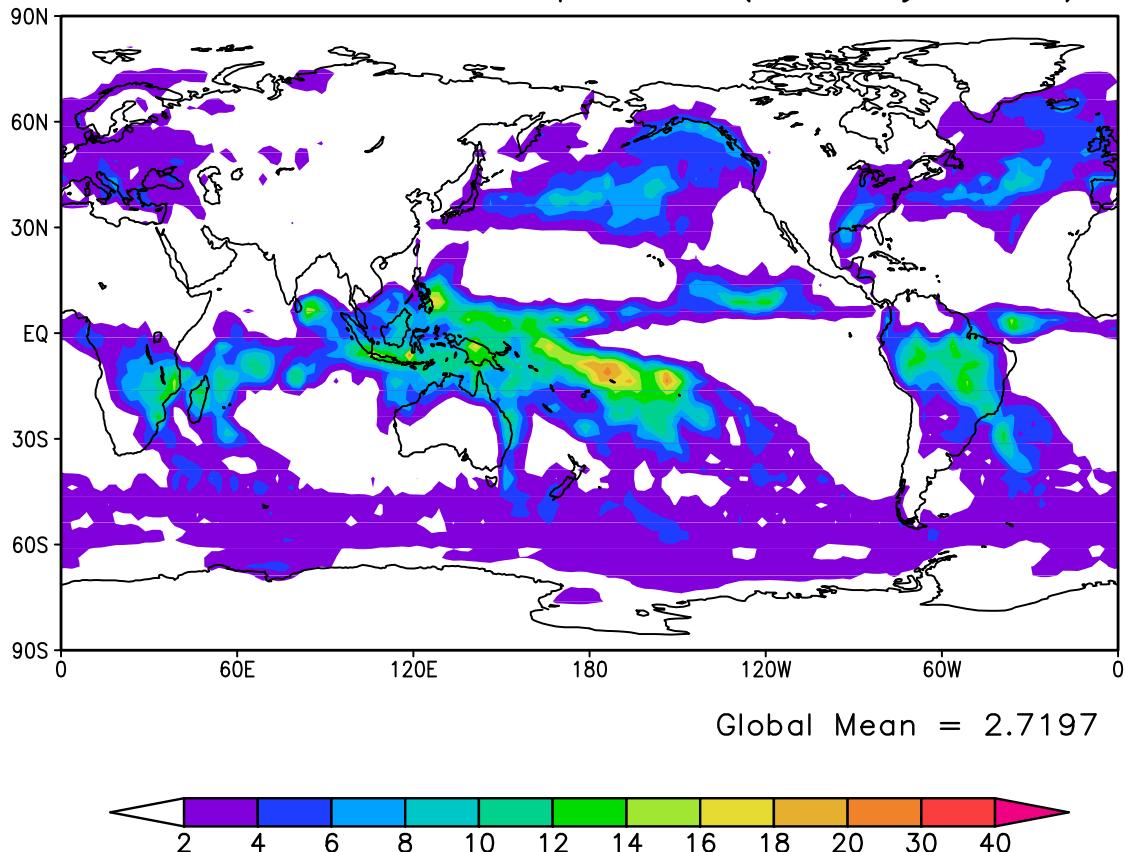


FIG. 7. Monthly mean precipitation from GPCPv2.2 2.5° valid for January 2013 at 12UTC (in mm/month).

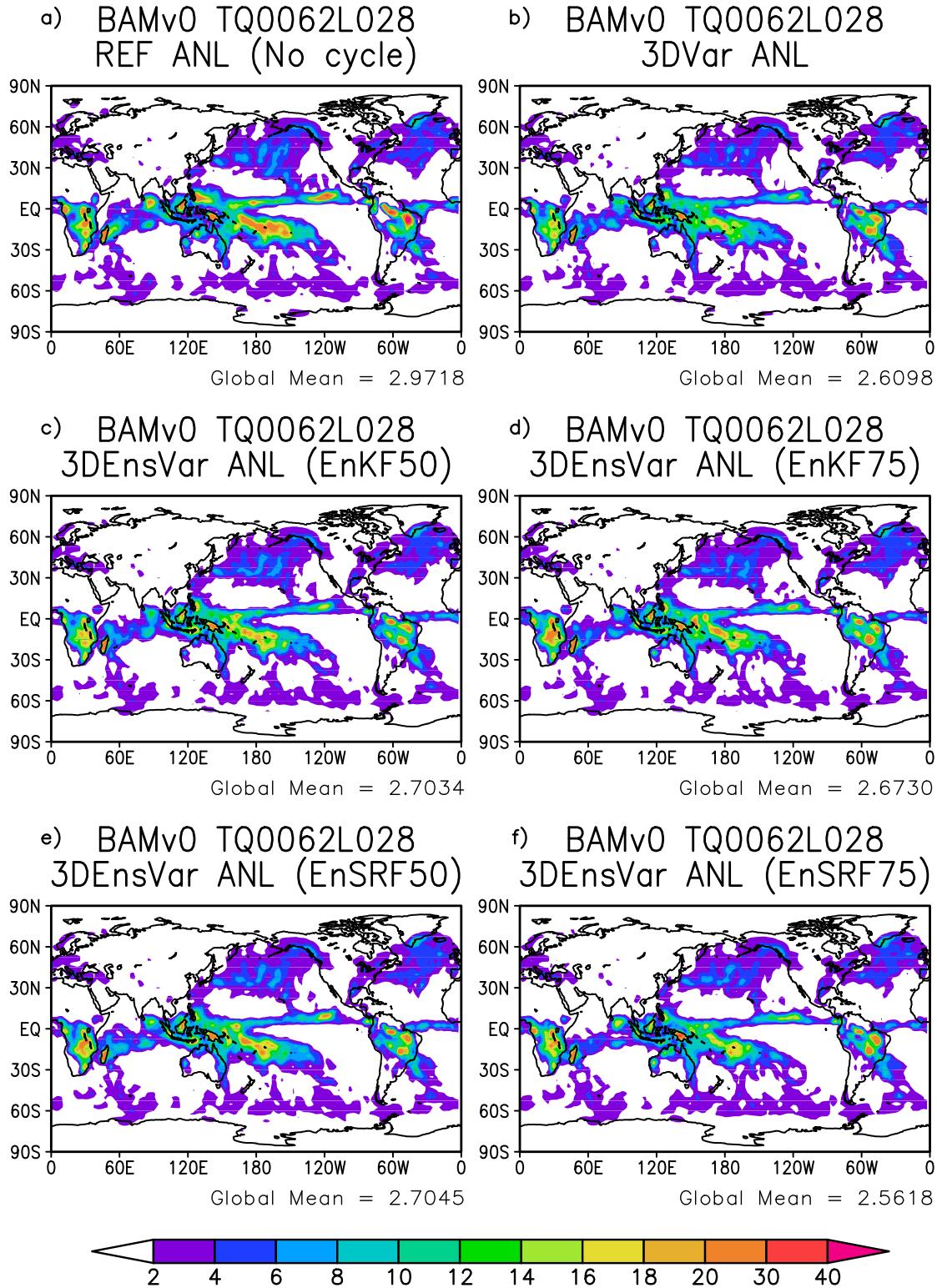


FIG. 8. Monthly average precipitation 24 hour forecast valid for January 2013 at 12UTC. Comparison is made between monthly mean precipitation from GPCPv2.2 and all experiments using different analyses. Below each picture, it is annotated the spatial average of the monthly mean of the precipitation forecasts, in mm/month.

Area Average 24 hr Prepic. 12z01Jan2013 12z31Jan2013

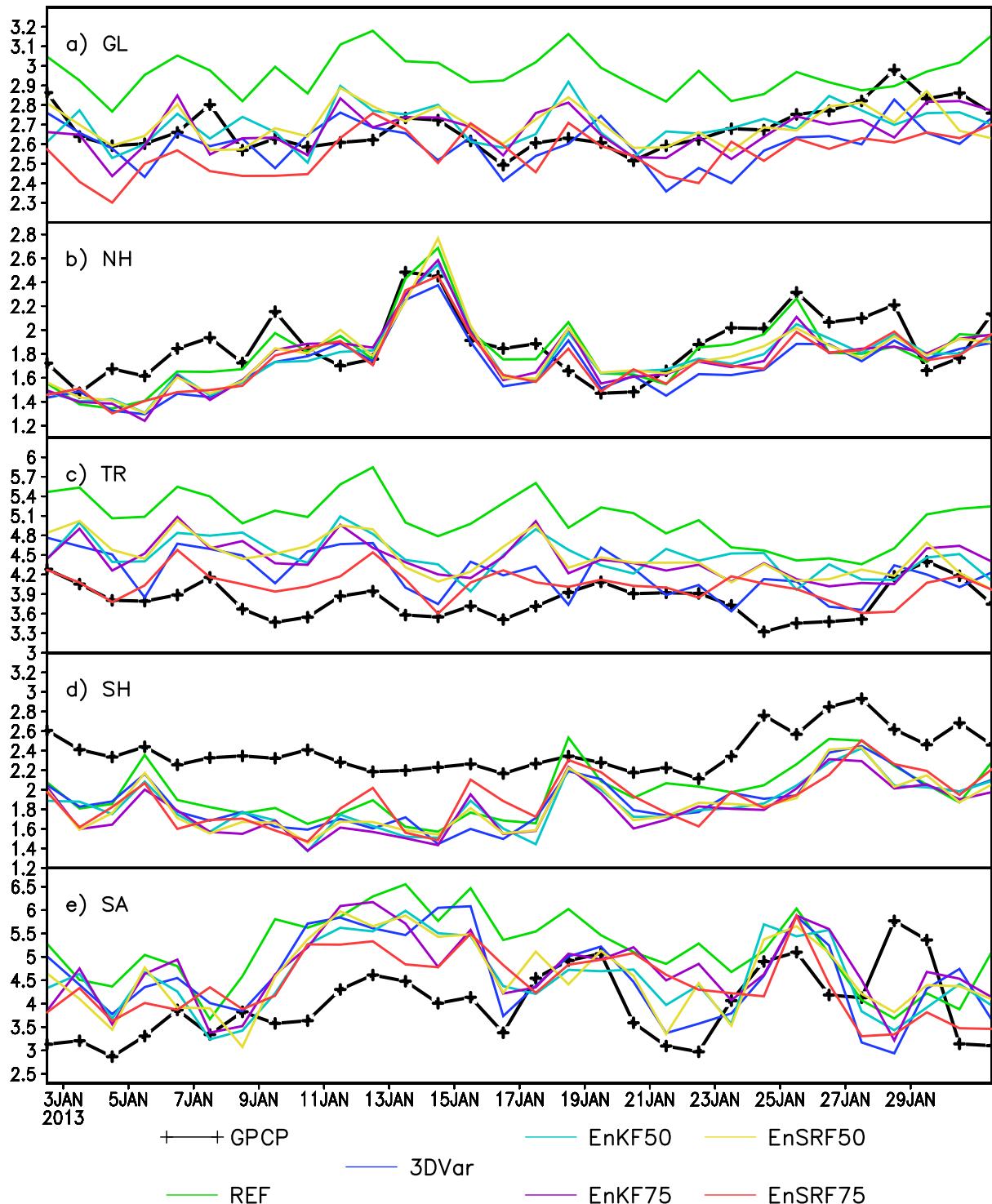


FIG. 9. Spatial averages of the 24 hour forecast precipitation (in mm/day) valid for January 2013 at 12Z. Panel a) precipitation for the whole globe (GL); b) same for the North Hemisphere (NH); c) for Tropics (TR); d) for South Hemisphere (SH) and e) for South America (SA). Precipitation from experiments are compared with the observed precipitation from GPCPv1.2.