

# 앙상블 예측(Ensemble Prediction)

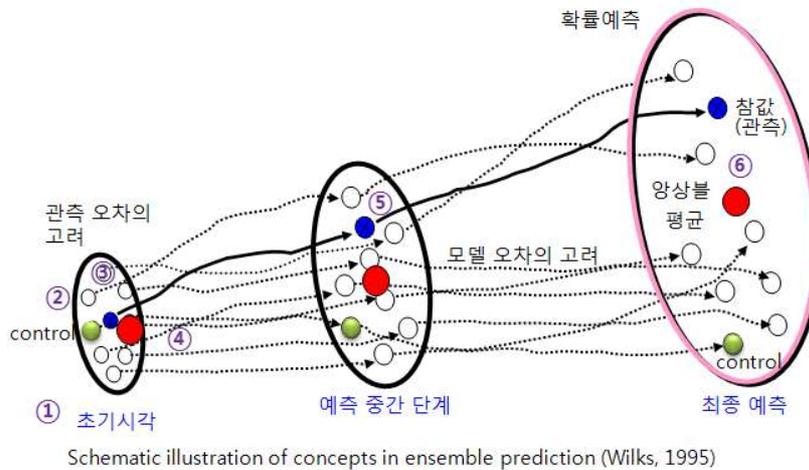
■ 발행: 예보국 ■ 문의: 예보기술팀(내선 1656/1657) ■ 발행일: 2012년 10월 8일(월)



## 앙상블 예측이란?

앙상블 예측이란 단일 수치예보가 가지는 결정론적인 예측의 한계를 보완한 것으로서 초기조건, 물리과정, 경계조건 등이 다른 여러 개의 모델을 수행하여, 확률적으로 미래를 예측하는 시스템이다. 앙상블 예측에는 기존 수치모델의 정보와 예보 불확실성에 대한 정보를 동시에 제공한다.

(a) 앙상블 예측 개념도



(b) 기상청 앙상블 예측 개념도 설명

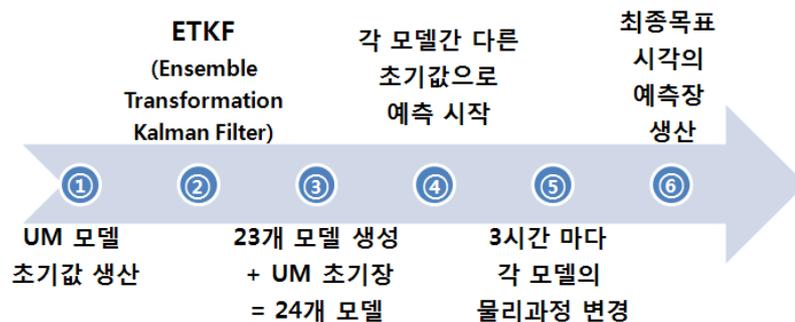


그림 1. 앙상블 예측 개념도와 요약 (a), (b)의 ①~⑥는 같은 의미임

그림 1은 앙상블 예측의 개념도와 UM 모델에서 생산하는 앙상블 예측에 대한 요약이다. 그림 1(b)를 보면, ① UM 모델의 초기장(N512 L70)이 생산되면, 이를 내삽하여 앙상블 초기장(N320 L70)을 생산한다. ② 섭동과정(UM 모델은 ETKF를 사용)을 통해, ③ 23개의 섭동 초기장이 생산된다. ④ UM 모델의 내삽된 초기장에 아무런 섭동을 추가하지 않고 모델을 수행하면 control 멤버가 되고, 초기장에 ETKF에 의한 분석장을 추가하여 모델을 수행하면 섭동멤버가 된다. 따라서, 총 24개의 앙상블 멤버들이 생산되어 예측을 시작한다. ⑤ 섭동멤버들은 3시간 예측마다 물리과정이 추가되어 시작 지점의 위치(실제 관측값)부터 최종예측 결과까지 차이가 난다. ⑥ 모든 멤버는 +10일까지 예측을 실행한다. 따라서, 각 모델은 예측시간이 길어질수록 최종목표 시각의 예측 편차가 커지게 되는데, 이렇게 시작 시점보다 예측값의 범위가 퍼져 분포하는 것을 스프레드(spread)라 부른다. 그림 1 (a)를 보면, 초기시각 관측과의 오차가 고려되어 예측을 시작하던 모델들은 최종 목표 시각에 도달할 때 참 값과 오차가 커지는데, 이 결과는 예측모델의 불확실성을 의미하는 것으로서, 단순히 하나의 모델에 의존해 예측을 할 경우 큰 오차가 발생할 수 있음을 보여준다. 이런 큰 오차발생을 줄이기 위해 각 모델 결과를 평균한 ‘앙상블 평균’값을 예측에 활용하고 있다. 표 1은 앙상블에서 자주 사용하는 용어에 대한 요약이다. 초기섭동 생성에는 Singular Vector, Breeding Method, Ensemble Transformation Kalman Filter 방법 등이 사용되고 있다. 기상청에서는 ETKF 방법을 사용한다.

표 1. 앙상블에서 사용하는 기본 용어

용 어	내 용
앙상블 예측 (ensemble prediction)	동일 예보시점에서 서로 다른 예보장들의 총체적 집합
멤 버 (member)	앙상블 예측에서 각각의 예보장들
섭 동 (perturbation)	모델과 관측과의 오차를 반영하여 모델분석장 또는 정상적인 물리과정의 변화를 계산한 후에 그 위에 추가로 부가하는 작은 값들
컨트롤(control)	초기 분석장에 섭동을 주지 않은 예보장
앙상블 평균 (ensemble mean)	전체 앙상블 멤버의 평균
스프레드 (Spread or "Uncertainty")	앙상블 평균에 대한 각 멤버의 표준편차

## 앙상블 예측의 장단점

표 2. 앙상블 예측의 장·단점 요약

요 소	장 점	단 점
결과물	단일 전지구 예측모델보다 우수, 예측 2일 이후, 중기 예측에 활용도 높음	경제성 때문에 낮은 해상도를 사용하며, 평활화(smoothing)로 중규모 기상파악이 어려움. 자료량이 방대하여 편차가 클 경우 해석의 난이도 증가
위험기상	발생 가능한 다양한 위험기상 시나리오 제공	확률정보의 해석 능력이 바탕이 되어야 함
비용, 시설	기상요소별 확률적 예측정보 제공	고성능의 슈퍼컴퓨터 필요

표 2는 앙상블 예측의 장·단점을 요약한 것이다. 수치모델의 오차는 시·공간적으로 차이가 있지만 일반적으로 예보기간이 길어지면 그 불확실성의 정도가 현저하게 커진다. ① 전지구 예측모델의 경우 예측 2~3일 이후는 단일 모델의 예측 신뢰도가 떨어지는데, 앙상블예측은 보다 긴 선행시간에서의 예측성을 향상시켜 중기예보에 활용도가 높다. ② 앙상블예측은 긴 선행시간에 대한 예측성의 향상뿐만 아니라, 예측의 불확실성(또는 신뢰도)에 대한 정보도 같이 제공한다. 따라서 예보관은 불확실성에 대한 정보를 가지고 상황에 맞는 정보생산과 예보결정을 내릴 수 있다. ③ 앙상블예측은 또한 가장 가능성 있는 하나의 예측이 아니라 앙상블이 표현해 내는 범위에서 가능한 모든 예측시나리오를 제공하므로, 예측기간 동안에 재해기상을 유발할 수 있는 기상현상의 발생가능성과 어느 정도의 강도를 가지는 재해기상의 발생이 가능한지에 대한 정보도 얻을 수 있다. ④ 그렇지만 전지구 모델 하나를 계산하는데도 엄청난 계산량이 요구되는데, 동시에 수십 개의 모델을 수행해야하는 앙상블예측시스템은 독자적인 슈퍼컴퓨터를 보유하고 있지 않는 기관에서는 운영하기 힘들다. 이런 이유로 전 세계에서 10개 기관만 전지구 앙상블 모델을 현업으로 운영하고 있다.

## 앙상블 산출물의 활용법

앙상블 예측에서 각 멤버들의 편차인 스프레드의 이해가 중요하다. 표 3은 스프레드가 큰 경우와 작은 경우에 대한 의미와 해석, 주의사항을 요약한 것이다.

표 3. 앙상블 스프레드 해석 요약

	스프레드가 작은 경우	스프레드가 큰 경우
의 미	모델의 예측성이 높다	모델의 예측성이 낮다
해 석	상세한 예보정보, 앙상블 평균 또는 중앙값(메디안)을 이용한 예보	상세한 예보를 피하고, 관측과 비교를 통해 가장 가능성이 있는 멤버를 감안하되, ECMWF, UM 모델의 앙상블 비교와 예측경향을 파악하고 다양한 시나리오 검토
주 의	종관규모 예측성이 좋다는 것이 지면변수(온도와 적운 강수 등) 예측성도 좋다는 것을 의미하지는 않음	업데이트되는 예보에 주의를 기울일 것 앙상블 멤버 중 최악의 시나리오 대비

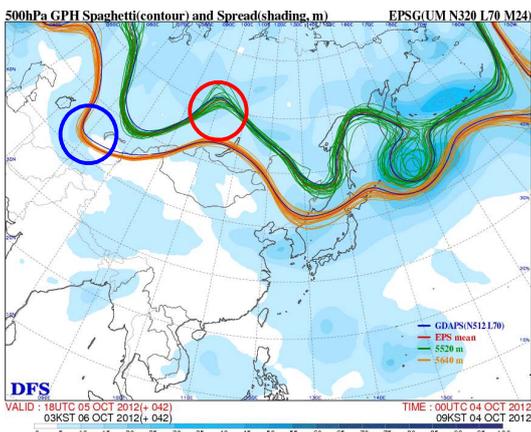


그림 2. 500hPa 고도상의 스파게티와 스프레드(빨간색 원은 파란색 원에 비해 스프레드가 크다. 즉, 불확실성이 크다.)

# 1) EPSgram

EPSgram은 지점·기상요소별 중기예보 결정과 예보의 신뢰도 제공에 활용할 수 있다. 주요 도시의 전운량(하늘을 10등분 한 것 중 구름이 덮고 있는 비율, ECMWF는 8분할 단위), 6시간 누적 강수량, 지상 풍속(지상으로부터 10m 고도로 내삽한 바람의 풍속, 단위 m/s), 지상 기온(지상으로부터 2m 고도로 내삽한 기온)을 제공한다. 앙상블 멤버의 예측값을 크기 순으로 나열하였을 때 아래부터 최소값, 10%, 25%, median(50%), 75%, 90%, 최대값의 크기를 보여준다.

붉은색 파선은 앙상블 멤버들의 평균 값이고, 파란색 실선은 GDAPS 결과이다. 예측시간이 길수록 그래프의 편차가 커지게 되는데 주변의 예측 시간대에 비해 편차가 클 경우 모델의 예측불확실성이 크다고 할 수 있다. 전운량, 강수량의 경우 대부분 앙상블 멤버가 0이더라도 0이 아닌 멤버가 있을 경우 평균값(붉은색 파선)이 확률분포의 중앙(median) 값보다 큰 값으로 표출된다.

연직시계열도는 모델 격자값을 관측지점으로 내삽한 값이 아닌, 관측점에서 가장 가까운 격자점의 값을 그대로 사용한다. 따라서 EPSgram의 해석은 모델 격자점이 관측과 어떤 위치에 놓여있는지를 감안하여 예측결과를 해석하여야 한다. 그림 4는 현업앙상블 EPSgram을 제공하는 주요도시에 대해 관측지점과 모델 격자의 위치를 표시한 것이다. 해안선과의 거리, 지형고도 등의 차이로 실제관측점과 모델 격자점의 특성에 차이가 날 수도 있다. 반드시 강수확률 분포 등의 수평분포 그림도 같이 참조해야 한다.

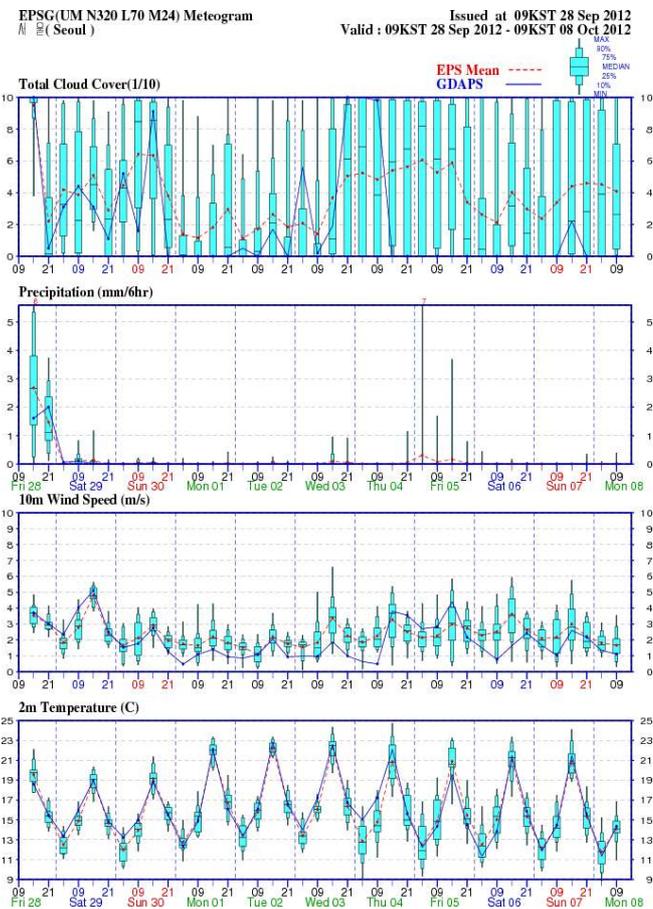


그림 3. EPSgram

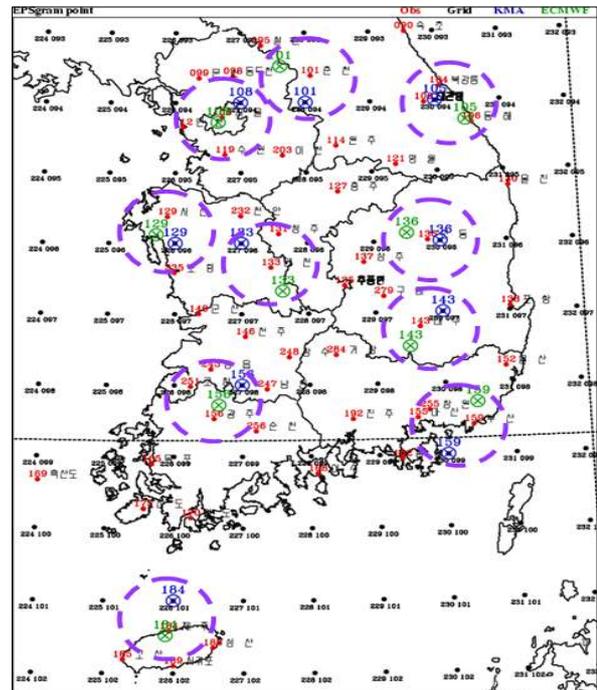


그림 4. EPSgram을 표출하는 주요 도시에 대해 관측지점과 모델격자 지점의 위치 비교. 빨간 점은 관측 지점, 파란 x는 현업앙상블에서 제공하는 EPSgram의 격자점, 초록 x는 ECMWF EPSgram의 격자 위치임

## 2) 시계열 예측경향

시계열 예측경향은 시간지연 앙상블(time-lag ensemble) 개념을 도입하여 앙상블 멤버의 개수를 늘리고, 수치예보 초기시각에 따라 예보가 변화하는 경향을 한눈에 파악할 수 있도록 표출한 것이다. UM EPSgram의 경우 동일 예보시점에 대해 지난 2.5일 동안 앙상블이 예측한 6개의 시계열을 시간순으로 제공한다. 가장 최신 자료가 위에 놓이며, 날짜의 위치(x축)가 동일한 날로 고정되어 있어 예보관이 모델 예측경향을 쉽게 파악할 수 있다. 또한 분석 시각 기준으로 2.5일 전까지는 관측자료가 이미 있으므로, 각 지점의 해당변수에 대한 최근의 앙상블 예측 오차 정도 및 경향을 파악하고 앞으로 다가올 기간에 대한 앙상블 예측을 해석하는데 참고자료로 활용 가능하다. 지난 기간에 대한 앙상블 예측과 관측을 검은 점으로 함께 표출하고 있다(그림 5 참조). 예보관은 각 모델의 예측 일관성을 파악할 수 있으며, 큰 차이 없이 비슷한 예측 분포를 보일 경우 예보결정에 활용 가능하며, 일관성이 결여된 경우 관측과 모델간의 차이, 6개의 모델 중 관측과의 오차가 가장 적은 모델의 선택 등 다양한 접근 방법으로 예보결정에 활용해야 한다. 제공변수로는 6시간 누적 강수, 2m 기온이며, +10일 까지 예측 자료를 제공한다. ECMWF EPSgram의 경우 동일 예보시점에 대해 지난 2일 동안 앙상블이 예측한 5개의 시계열을 시간순으로 한 그림에 표출하며 제공 변수는 강수, 운량, 풍속, 기온이다.

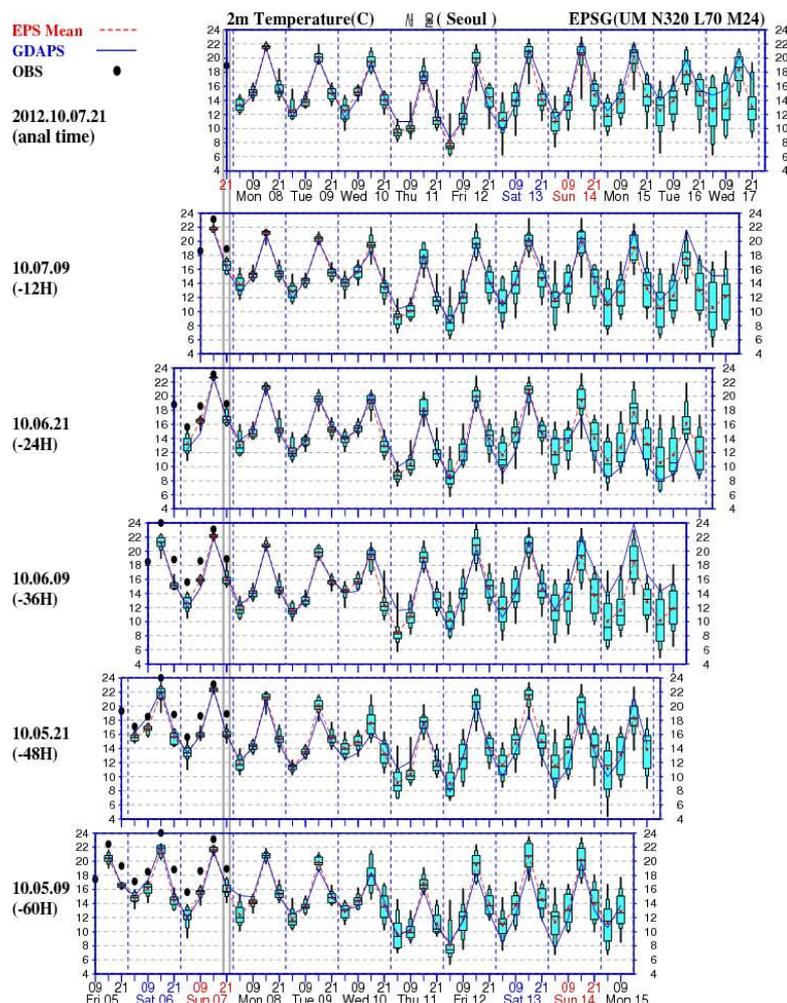


그림 5. UM-EPS 기온시계열 예측경향

### 3) 극값예측지수 (EFI: Extreme Forecast Index %)

극값 예측지수는 앙상블 예보의 분포가 모델의 기후<sup>1)</sup> 분포에서 얼마나 벗어나는지를 나타내는 지표이다. 모델 기후에서 크게 벗어나는 값은 계절과 기상요소에 따라 위험기상 발생 가능성을 암시하는 것이므로 수일 전에 위험기상 발생가능성을 인지할 수 있는 경보상황인지 자료라고 할 수 있다. 현재 앙상블 예측이 모델기후의 확률분포에서 벗어난 정도를 지수로 표현하여 각종 재해기상(호우, 강풍, 한파 및 폭염) 가능성의 정보를 백분율로 제공한다. 이 때, 극값예측지수는 -100에서 100 사이의 값을 가지며 0에 가까울수록 모델기후자료 분포와 유사함을 의미하고 백분율의 절대값이 클수록 기후분포에서 벗어난 것이므로 재해기상 가능성이 높음을 보여준다. 각 기상요소별 확률분포 예상이 -40% 이하, +40%이상일 경우 녹색, 붉은색 계열로 채색되어 있다(그림 6).

일 강수량은 기후보다 약한 강수는 의미가 없으므로 호우의 가능성을 나타내는 양의 값만 표출하며 값이 클수록 모델기후 산출기간에 예측했던 과거의 강수량에 비해 많은 양이 예측됨을 의미한다. 일 최대풍속의 경우도 강수와 마찬가지로 양의 값만 표출하며 값이 클수록 강풍의 가능성이 큼을 의미한다. 일 최저, 최고기온은 음으로 커지면 모델기후보다 춥고 양으로 커지면 모델기후에 비해 따뜻함을 의미한다. 바탕에 파란색 실선은 모델 기후평균이다. EFI는 모델기후의 차이임을 명심해야 한다. 그러므로 동일한 EFI 값을 보이더라도 계절에 따라 발생하는 기상현상의 강도는 큰 차이가 날 수 있다.

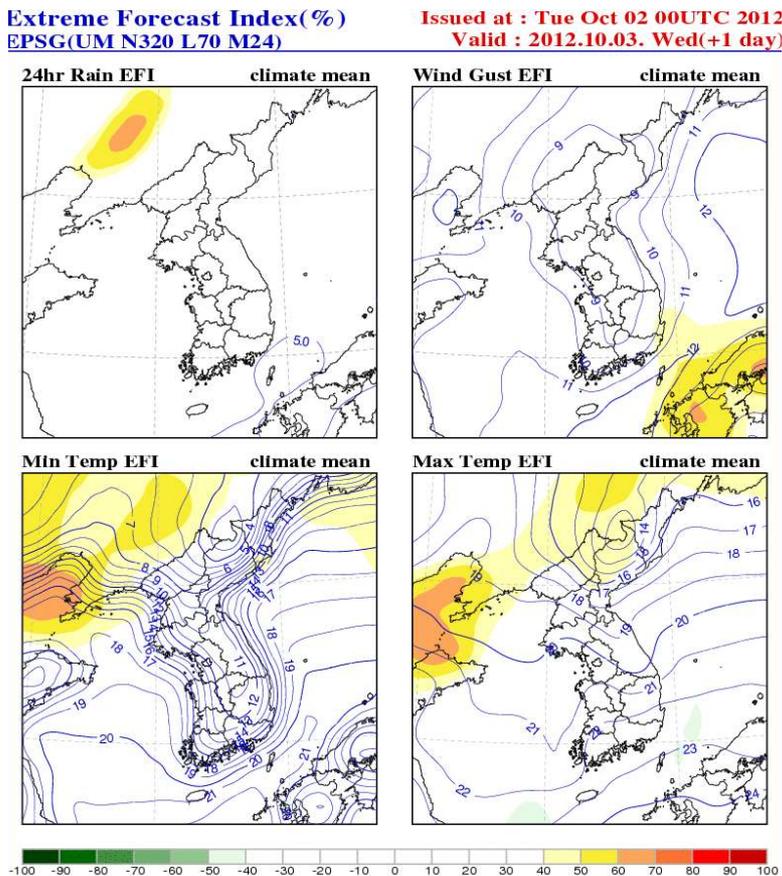


그림 6. 극값 예측지수 4종

1) 모델기후는 3년(2009~2011)동안 24개 앙상블 멤버에 대해 일 자료를 추출하여 해당 일의 앞, 뒤 15일 이동 창(window)을 가지고 산출한다.

## 4) 평균/편차

평균/편차도는 앙상블 멤버 23개와 Control 1개로 구한 앙상블 평균(contour)과 그 평균에 대한 멤버들의 표준편차를 스프레드(채색)로 나타낸 것이다(그림 7). 앙상블은 동일 예보시점에 대해 수십 개의 예측을 제공하므로 활용의 편의를 위해 가장 가능성이 있는 시나리오라고 볼 수 있는 앙상블 평균과 각 멤버들이 평균으로부터 얼마나 다르게 예보했는가를 나타내는 편차(스프레드)를 한 장에 표출하여 제공한다. 스프레드가 큰 지역은 예측의 불확실성이 크고, 스프레드가 낮은 지역은 예측의 불확실성이 적음을 의미한다. 제공하는 자료로서, 그림 6과 같이 동아시아 해면기압: 앙상블 평균(등치선, 2hPa 간격)과 편차(음영, 0.5hPa 간격), 동아시아 850hPa 기온: 앙상블 평균(등치선, 3°C 간격)과 편차(음영, 0.5°C 간격), 동아시아 500hPa 고도: 앙상블 평균(등치선, 60m 간격)과 편차(음영, 5m 간격)가 있다. 예보관은 기압계의 이동속도 및 기압골의 강도에 대한 모델의 불확실성을 참고하여 중기예보 등에 활용할 수 있다.

## 5) 스파게티

스파게티는 모든 앙상블 멤버의 예측을 하나의 그림에 그리면 매우 복잡하므로 원하는 등치선만을 선택하여 각 멤버의 예측을 모두 보여주는 그림이다. 500hPa 특정고도 등치선(2쌍)을 한 장의 그림에 스프레드(음영(파란색 계열), 5m 간격)와 함께 표출한다(그림 8). 계절변화를 고려하기 위해 등고선 값을 달리하여 5340m부터 5880m까지 60m간격으로 시작하는 8쌍을 제공하며, 한 그림에서 두 등치선은 120m 간격으로서 총 8개로 구성되어 있다. 여름철 북태평양 고기압은 5880선을 참고할 수 있으며, 상층과동의 움직임은 이보다 낮은 등치선 값을 참고할 수 있다. 멤버간의 간격이 좁은 지역(스프레드 작음)은 일반적으로 모델의 예측성이 높다고 할 수 있으며, 멤버간의 간격이 넓은 지역은 일반적으로 예측 정확성이 낮고 스프레드가 크게 나타나는 것이 특징이다. 앙상블 스프레드를 함께 표출하고 있으므로 음영으로 표출된 지역과 참고하여 활용해야 한다.

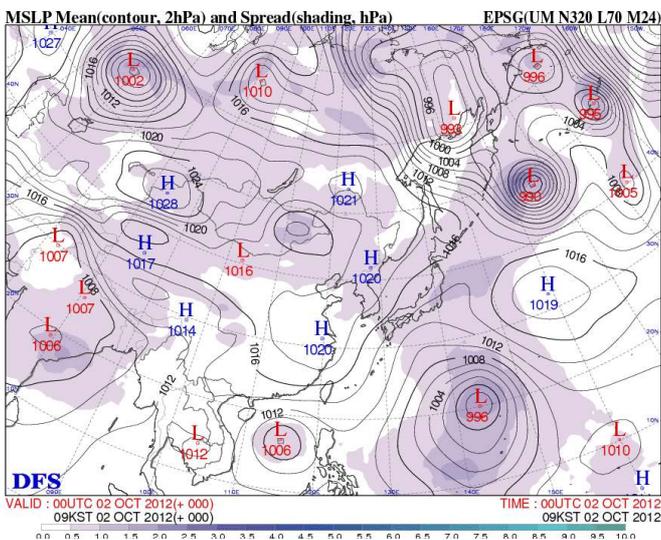


그림 7. 평균/편차도 예제

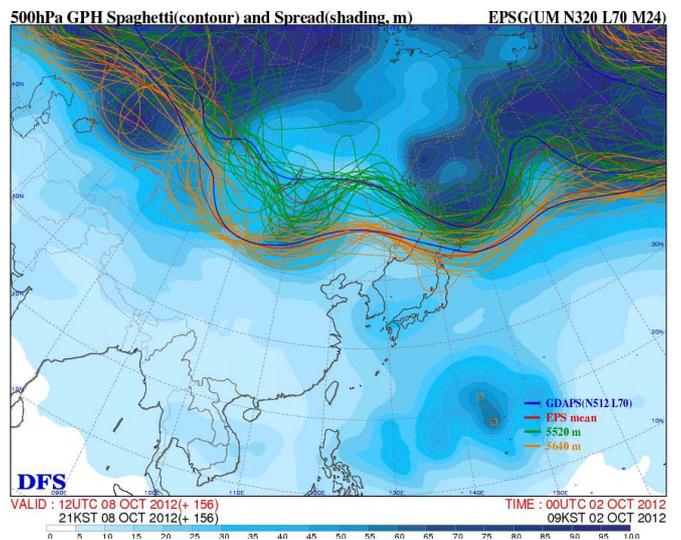


그림 8. 스파게티 차트 예제

## 6) 강수확률

강수확률은 6시간 누적 강수량을 4개의 계급으로 나누어, 앙상블 멤버들의 예보가 각 계급에 해당하는 확률을 정량적으로 계산하여 확률의 분포(shading)를 나타낸 분포도이다. 확률(P)의 계산은 격자마다 각 계급에 해당하는 강수량을 예측한 앙상블 멤버의 수를 M이라 할 때,

$$P = (M/24) \times 100$$

강수량계급: 1, 5, 10, 25mm / 6시간

그림 9에서 등치선은 해면기압의 앙상블 평균값이다. 일반적으로 가장 약한 강수에 대한 강수확률 예측을 먼저 확인해보고, 강수가능성이 있을 경우 어느 정도 강도의 강수가 예측되는지에 대한 정보를 얻기 위해 강수량 계급이 더 높은 구간에 대한 강수확률 분포도를 참고한다. 총 4개(1, 5, 10, 25mm/6시간)의 계급별로 메뉴가 구성되어 있으며, 각 단계별 예측한 앙상블 멤버가 전체 대비 몇 %인지를 강수확률로 표현한 것이다. 강수확률 50% 이상은 노란색, 70% 이상은 주황색으로 알아보기 쉽게 표현되어 있다.

## 7) 스탬프 맵(stamp map)

앙상블 예보는 평균장과 편차장을 제공하는 것도 중요하지만 모든 멤버의 시나리오를 함께 표출하여 예보에 활용하는 것도 중요하다. 스탬프 맵은 그림 10처럼 멤버들을 한 화면에 우표처럼 배열하여 표출한 그래프이다. 모든 멤버의 시나리오를 제공하여 앙상블 멤버들 중 극한기상(extreme weather) 시나리오에 대한 가능성을 검토할 수 있다.

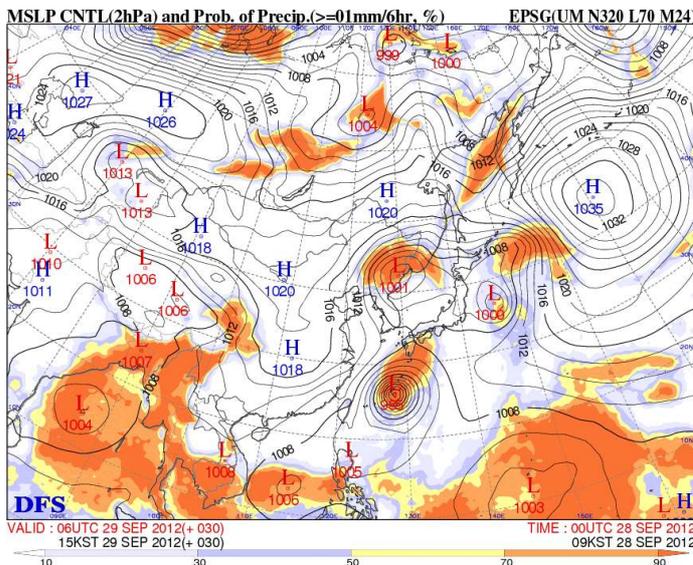


그림 9. 강수확률 차트

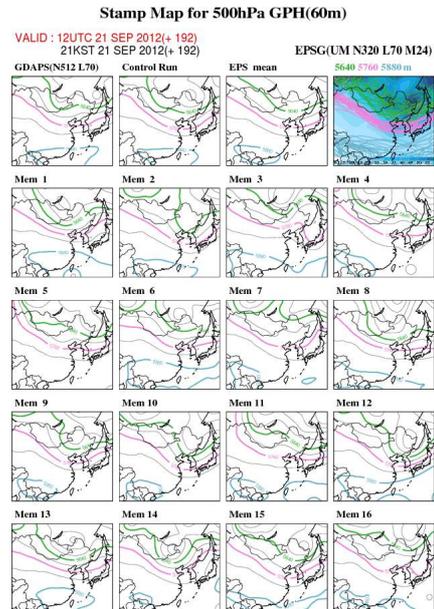


그림 10. 스탬프 맵 예제

<참고문헌>

김동준, 2007: Recent developments in KMA's ensemble prediction system(EPS), Geo-nwp training workshop 발표자료.

기상청, 2011: 수치예보 전문기술인력 양성과정 표준교재, 수치예보의 이해, 84~101pp.

손주형, 2012: 지경노 발표자료, 극값예측지수(EFI)와 앙상블개선.

Introduction of Ensemble Prediction, COMET module, [www.meted.ucar.edu](http://www.meted.ucar.edu)

Wilks, D. S., 1995: Statistical methods in the atmospheric science: An introduction academic press, 467pp.