

CMIP6 대응 새로운 기후변화 시나리오의 국제 데이터노드 구축

이재희^{†*} · 성현민^{**} · 심성보^{**} · 부경은^{***}

^{*}국립기상과학원 미래기반연구부 연구원
^{**}국립기상과학원 미래기반연구부 연구사, ^{***}국립기상과학원 현업운영개발부 연구관

International Data Node System for CMIP6 Climate Change Projections

Lee, Jae-Hee^{†*} · Sung, Hyun Min^{**} · Shim, Sungbo^{**} and Boo, Kyung-On^{***}

^{*}Researcher, Innovative Meteorological Research Department, National Institute of Meteorological Sciences(NIMS), Jeju, Korea
^{**}Research Scientist, Innovative Meteorological Research Department, NIMS, Jeju, Korea
^{***}Senior Research Scientist, Operational Systems Development Department, NIMS, Jeju, Korea

ABSTRACT

The National Institute of Meteorological Sciences (NIMS) of the Korea Meteorological Administration (KMA) is participating in CMIP6 to produce climate change projections for the IPCC Sixth Assessment Report (AR6). Based on the KMA Advanced Community Earth-system model (K-ACE) and UK Earth System Model (UKESM1) in collaboration with the MetOffice, UK, we will take part in ScenarioMIP and AerChemMIP to model long-term climate simulations under the Shared Socioeconomic Pathways for IPCC AR6. The experiment output dataset will be enormous, with management and world-wide dissemination critical to meeting global goals. In order to support data access, a node in the Earth System Grid Federation has been developed.

We applied the postprocessing method to publish CMIP6 data to the ESGF; this postprocessing method includes the international standard data conversion (CMORization) and quality assurance (QA) processes. These CMORization and QA processes are mandatory for publishing data to ESGF and increase the reliability of the data. We expect that shortening the time needed from model simulation to delivering data through the ESGF node will improve climate change science outcomes and will facilitate the CMIP activities of various users.

Keyword : Climate change, ESGF, Quality Assurance, CMOR, CMIP6

1. 서론

기후변화에 따른 사회경제적 영향 및 미래 국제협약에 포함될 수 있는 잠재적 대응 요소들에 대한 검토와 권고사항을 준비하기 위해 1988년 유엔환경계획(United Nations Environment Programme, UNEP)과 세계기상기구(World Meteorological Organization, WMO)에서는 기후변화에 관한 정부 간 협의체(Intergovernmental Panel on Climate Change, 이후 IPCC)를 설립하였다. IPCC에서는 온실가스

배출에 의한 잠재적인 미래 기후변화 시나리오를 산출하여 평가보고서에 활용한다(IPCC, 2013). IPCC 평가보고서는 기후위기에 대응하는 여러 정책 결정을 위한 중요한 도구로 사용되어 기후변화 대응 전략을 평가한다. 따라서 많은 국가들이 IPCC 평가보고서 작성을 위한 기후변화 시나리오 개발에 참여하고 있다. 시나리오는 세계기후연구프로그램(World Climate Research Programme, 이후 WCRP)의 결합 모델 상호비교 프로젝트(Coupled Model Intercomparison project, 이후 CMIP)를 통해 산출된다. 실험을 설계하고 실

[†]Corresponding Author : jhlee0@korea.kr (33 Seohobuk-ro, Seogwipo-si, Jeju 63568, Korea, National Institute of Meteorological Sciences, Tel. +82-64-780-6640)

ORCID 이재희 0000-0003-3227-923X 심성보 0000-0002-3533-5818
성현민 0000-0001-5173-8580 부경은 0000-0002-1505-578X

행하며 모델 출력을 생산하는 전문가 외에도 기후 변화를 연구하는 관련 분야의 연구자, 기후 서비스를 개발하고 제공하는 개발자 등 많은 사람들이 CMIP 산출물을 활용하게 된다.

자료는 생산하는 만큼 사용자에게 제공되는 것도 매우 중요하므로 모델링 커뮤니티의 전문가가 아닌 일반 이용자에게도 자료의 투명성과 가용성이 보장되어야 한다. 그리고 데이터 생산 후 오랜 시간이 지나도 모델의 구성, 생산 과정의 절차와 선택에 대한 사항까지 추적할 수 있어야 한다. 이러한 요구에 의해 2014년 WCRP의 기후 모델링 워킹 그룹의 기반시스템 패널(WGCM Infrastructure Panel, 이하 WIP)이 설립되어 CMIP을 지원하는 업무를 담당하고 있다 (Eyring et al., 2016a).

WIP는 CMIP 모델 출력에 대한 이용과 분석을 용이하게 하는 소프트웨어 개발 프로젝트를 지원하고 있다. 모델 출력이 공통의 구조와 형식으로 작성되도록 하고 공통의 방법으로 전 세계적으로 보관 및 접근 가능하게 하는 표준을 정의한다. 또한 CMIP 또는 다른 MIP 프로젝트가 과학적 요구를 충족시킬 수 있도록 CMIP 실험의 설계를 감독하고 권고 사항을 제공한다. 그 외 WIP에서는 CMIP을 뒷받침하기 위한 계산 기능과 자료 기반 시스템을 조사하며, 국제표준체계의 실험 설계를 감독하는 팀과 자원을 제공하고 인프라를 설계하는 팀 간의 의사소통을 개선하는 업무를 수행한다. 특히 WIP는 CMIP 설계, 미래 성장을 지원하는 데 필요한 글로벌 데이터 인프라(Earth System Grid Federation, 이후 ESGF) (Williams et al., 2011a, Williams et al., 2011b)에 대한 권고사항을 제공하는데, 이 권고안에 따라 각 모델링 센터는 ESGF 노드를 구축하고 운영하여 CMIP 자료 공유를 제공한다.

기상청 국립기상과학원에서는 2007년부터 기후변화시나리오 산출을 통하여 IPCC 4차, 5차 평가보고서에 기여하여 왔다. 2022년 IPCC 6차 평가보고서에 대해서도 CMIP6에 참여하여 기후변화 시나리오를 산출 중이다. 국제적으로 5차 평가보고서에 비해 6차 평가보고서는 모델의 향상, 기후 과정 진단을 위한 다양한 실험의 증가 등으로 인해 자료의 용량은 크게 증가하였다. 이에 기관별 생산된 자료들이 수요자들에게 대용량으로 처리, 배포, 관리하는 자료 유통, 전달이 매우 중요해진다. 국립기상과학원에서는 자료의 산출에서 제공까지의 시간을 단축하여 자료제공을 신속히 하고 시나리오의 활용을 높이고자 ESGF 노드를 구축하였는데 본 논문에서는 이 구축과정을 소개하고자 한다. WIP의 권고안을 준수한 국제 데이터노드 구축과정과 자료 제출 과정에 대한 내용을 포함하였다. 2장에서는 CMIP6에 참여하여

생산한 자료를 언급하고, 3장에서는 ESGF 자료 제출에 필요한 과정인 국제표준자료 변환, 품질 보증과정에 대하여 설명하고, ESGF의 역할, 구축 방안 및 CMIP6 자료 제출에 대하여 논의한다. 마지막으로 4장에서는 요약 및 결론을 제시하였다.

2. CMIP6 전지구 기후변화 시나리오

2.1 CMIP6 대응을 위한 기후변화 시나리오 자료 산출

온실 가스나 에어로졸과 같은 인위적 강제력과 태양 활동과 화산폭발 같은 자연 강제력이 기후변화에 미치는 영향을 살펴보고, 이러한 외부강제력 변화에 따른 기후 전망정보를 생산하기 위해서는 기후 모델링이 뒷받침되어야 한다. 이에 기후모델을 활용한 미래 시나리오 산출과 불확실성을 평가하는 방법론이 제시되고, CMIP의 기초가 되었다. IPCC의 첫 번째 보고서에서 공식화 되어진 대기 모델 비교 프로젝트(Atmospheric Model Intercomparison Project, 이후 AMIP)는 참여를 원하는 누구에게 열린 최초의 모델간 비교 프로젝트이다(Gates, 1992). 이때 5개의 모델이 참여하였고, AMIP, 산업화이전 제어실험, abrupt 4×CO₂, 1% CO₂ 로 구성된 DECK(Diagnostic, Evaluation and Characterization of Klima) 실험이 표준화되었다(Balaji et al., 2018). 이렇게 시작된 CMIP은 현재 CMIP6에 이르러서 40여개의 모델링 센터에서 100개 이상의 모델을 가지고 CMIP 패널에 의해 승인된 23개의 MIP에 참여하고 있다. CMIP6의 MIP에서는 CMIP5에 비하여 상당히 확장된 287개의 실험이 요구되고 있다. 287개 실험의 제출 변수 목록에는 8개의 모델링 영역(대기, 해양, 지면, 해빙 등)에서 산출되어야 할 수천 개의 변수가 존재한다(<https://earthsystemcog.org/projects/wip/CMIP6DataRequest>, accessed 2020 May 30). 또한 물리, 화학, 생물학적 프로세스의 영향으로 실험에 대한 복잡도가 증가되어 자료의 용량이 크게 증가하였다. 국립기상과학원에서는 필수 실험인 DECK을 비롯하여 SSP에 따른 시나리오 모델 상호비교 프로젝트(Scenario Model Intercomparison Project, 이후 ScenarioMIP)와 에어로졸 화학 모델 상호비교 프로젝트(Aerosol Chemistry Model Intercomparison Project, 이후 AerChemMIP)에 참여하여 자료를 산출하고, ESGF에서 제공 하고 있다.

2.2 새로운 온실가스 경로에 따른 미래 기후변화 시나리오

미래 기후변화에 대한 예측은 기후시스템에 대한 이해를 높이고, 사회적 위험과 대응 방안을 수립하는데 중요한 역할을 한다. IPCC AR6(Sixth Assessment Report)에서는 Fig. 1과 같은 새로운 온실가스 경로(Shared Socioeconomic Pathway, 이후

SSP)를 제안하고 ScenarioMIP을 이용한 미래 기후변화 시나리오를 산출을 요구하였다(Eyring et al., 2016a; O'Neill et al., 2014; O'Neill et al., 2016).

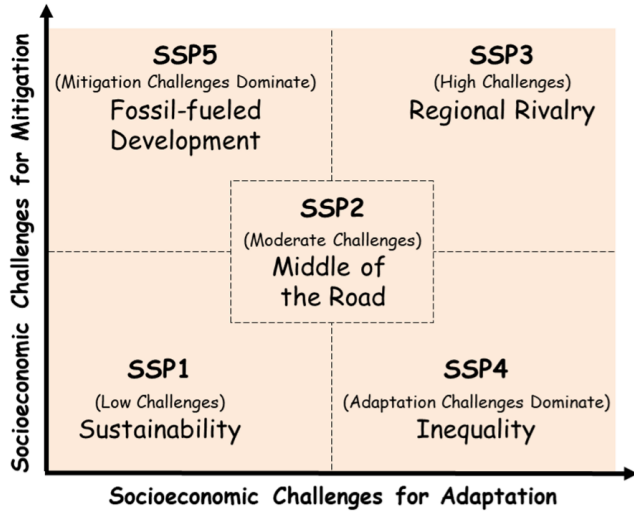


Fig. 1. The Shared Socioeconomic Pathways(SSP). It is divided into five domains by the challenges of mitigation and adaptation to climate change. (Fig. 1 in O'Neill et al. (2014))

CMIP6의 주요 활동에 해당하는 ScenarioMIP은 다음과 같은 목적을 가지고 실험이 수행된다. 첫째, 시나리오에서의 물리적 기후 시스템과 기후변화가 사회에 미치는 영향에 대한 이해도를 향상시켜 연구협업을 용이하게 하는 것이다. SSP-RCP 온실가스 농도경로는 통합평가모델(Integrated Assessment Model, IAM)을 활용하여 경제·사회적 요소(인구증가, 교통·식생활·에너지 분야의 변화, GDP 등)에 관한 미래전망을 고려한 사회경제 시나리오에 탄소순환이 결합되어 개발되었다. ScenarioMIP 실험의 결과는 여러 커뮤니티의 통합 연구를 촉진할 수 있도록 미래 시나리오를 위한 새로운 기후 정보를 제공할 것이다. 둘째, ScenarioMIP 및 다른 CMIP6 프로젝트의 시나리오기반 연구에서 강제력 측면의 기후 효과와 관련된 과학적 질문을 해결할 수 있는 근거를 제공하는 것이다. 마지막으로 다중 모델 앙상블에 기초한 예측 불확실성을 계량화하는 개선된 방법을 제공하는 것으로 모델 성능, 모델 의존성 및 관측 불확실성을 고려하는 것이다. 국립기상과학원에서는 ScenarioMIP의 SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5를 이용하여 미래 기후변화 시나리오를 산출하고 21세기말 평균 기온과 강수량, 해수면 온도, 해빙 면적, 극한 기후값을 예측하였다. 이에 사용된 기

후 모델은 국립기상과학원 지구시스템 모델인 K-ACE(Lee et al., 2019)와 영국기상청 지구시스템 모델인 UKESM1(Sellar et al., 2019)이다. 산출된 자료는 ESGF 노드에 제출되어 연구자들이 이용할 수 있도록 하였다. Table 1은 ScenarioMIP의 4가지 실험에 대하여 보여준다. SSP1-2.6 시나리오는 사회 불균형의 감소와 친환경 기술의 빠른 발달로 기후변화, 적응능력이 좋은 지속 성장가능 사회경제 구조의 저탄소 시나리오이고 SSP2-4.5 시나리오는 중도성장의 사회경제 시나리오로 기후변화 완화 및 사회경제 발전 정도가 중간 단계를 가정하는 시나리오이다. SSP3-7.0 시나리오는 사회경제 발전의 불균형과 제도적 제한으로 인해 기후변화에 취약한 상태에 놓이는 사회경제 구조의 시나리오이다. SSP5-8.5 시나리오는 기후정책 부재, 화석연료 기반 성장과 높은 인적 투자로 기후변화 적응 능력은 좋지만, 완화능력이 낮은 사회경제 구조의 고배출 시나리오이다. 파리협정에 의거한 1.5도 기온상승 억제와 가장 유사한 시나리오는 SSP1-2.6 시나리오이며, 현재와 같은 추세가 지속되는 미래와 유사한 시나리오는 SSP3-7.0 이다.

Table 1. ScenarioMIP experiments.

| Scenario | Description |
|----------|--|
| SSP1-2.6 | Scenario represents the low end of the range of future forcing pathways. It combines low vulnerability with low challenges for mitigation as well as a low forcing signal. |
| SSP2-4.5 | Scenario represents the medium part of the range of future forcing pathways. It combines intermediate societal vulnerability with an intermediate forcing level. |
| SSP3-7.0 | Scenario represents the medium to high end of the range of future forcing pathways. It combines relatively high societal vulnerability with relatively high forcing. |
| SSP5-8.5 | Scenario represents the high end of the range of future pathways. It is a scenario with the highest radiative forcing, because of high CO2 emissions. |

2.3 미세먼지와 대기화학에 대한 기후변화 시나리오

에어로졸 및 오존, 메탄 등 대기화학 성분에 대한 정책적 자료 지원을 위하여 국립기상과학원에서는 영국기상청과 협력하여 지구시스템 모델(UKESM1)을 이용하여 AerChemMIP 프로젝트(Collins et al., 2017)에 참여하였다.

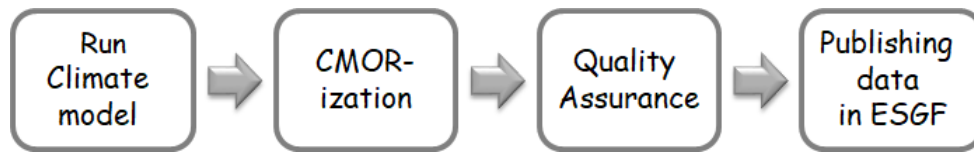


Fig. 2. The processes for Publishing CMIP6 data in ESGF. In order to submit CMIP6 data, the international standard data conversion (CMORization) and quality assurance(QA) is mandatory.

이 프로젝트는 에어로졸과 화학반응 가스가 기후변화에 미치는 영향을 이해하기 위해 황산염, 검댕, 유기탄소와 같은 에어로졸 성분뿐만 아니라 메탄, 오존, NOX, VOC 등 화학성분별 실험으로 구성되어 있다. 이는 과거기간동안의 인위적 배출이 전 지구 복사강제력과 지역 기후에 영향을 주었는가, 미래 기후 정책에 따른 근기 기후변화 유발물질(Near-Term Climate Forcers, 이후 NTCF)의 농도 변화 및 기후에 미치는 영향이 얼마나 되는지, 과거 NTCF 배출에 있어서의 불확실성이 복사 강제력 추정에 영향을 주는가, 자연적 NTCF 배출, 대기조성, 복사 효과에 기후 피드백이 얼마나 중요한가 등의 과학적 질문에 대한 해답을 얻는 것을 목적으로 하고 있다.

시나리오 산출에 사용된 영국기상청 지구시스템 모델(UKESM1)은 메탄, 오존, 질소 성분과 관련된 대기 화학반응과 에어로졸의 성분 별 물리 과정을 고려한다. 기존 CMIP5 모델과 비교하여 에어로졸의 생성 및 성장 뿐만 아니라 에어로졸-구름 상호작용에 대한 물리과정이 크게 향상되었다. 국립기상과학원에서는 AerChemMIP의 여러 실험 중 hist-piNTCF, histSST-piNTCF, piClim-NTCF, ssp370-lowNTCF 실험에 참여하여 자료를 생산하고 제출 중에 있다. 이러한 기후변화 시나리오 자료는 다양한 주제의 기후변화 연구에 활용되고 있다.

3. 기후변화 시나리오 자료 제출을 위한 ESGF 노드 구축

3.1 ESGF 국제 데이터노드의 필요성

CMIP에서의 산출물은 센서와 위성 등에 의해 생성되는 기상 관측 자료와 마찬가지로 전 세계 많은 연구자들을 위한 중요한 과학적 자원이다. 이러한 대용량의 데이터를 관리하고 공유하는 것 또한 자료 생산 못지않은 중요한 업무이다. 이를 위한 해결책으로 CMIP5에서는 ESGF를 구성하였다. ESGF는 데이터 보관 및 배포를 위한 기반시스템으로

2013년 세계 기후 연구프로그램의 공동과학위원회(WCRP Joint Scientific Committee)에 의해 제안되었다. CMIP3의 35TB에서 CMIP5는 3.5PB로 100배 증가된 데이터는 이러한 기반 시스템에 대한 관심을 증폭 시켰다(Balaji et al., 2018; Cinquini et al., 2014).

기후모델 실험을 설계하고 실행하며 결과 생산을 담당하는 지구 시스템 과학 전문가 외에도 건강, 농업, 천연 자원의 다양한 분야에서 기후 변화를 연구하는 연구자, 기후 서비스를 개발하고 제공하는 개발자 등 많은 사람들이 CMIP 자료에 관심을 갖고 있으며 이를 이용하고 있다. 이에 자료에 대한 과학적 재현성과 기초적인 내구성, 데이터의 출처에 대한 입증은 중요한 문제라고 할 수 있다. 따라서 ESGF 자료 제출을 위해서는 Fig. 2 와 같은 일련의 과정을 수행해야 한다. 먼저 후처리를 통하여 국제표준자료로의 변환(CMORization)을 수행한다. 변환된 자료는 자료가 가진 메타정보 및 데이터 검증을 수행하여 자료의 신뢰성을 높이고, 이후 ESGF 노드에 공개하여 다수의 연구자가 이용할 수 있도록 한다. 각 과정에 대한 세부적인 내용은 다음 절에서 설명한다.

3.2 CMIP6 기여자료 산출과 품질보증

3.2.1 국제 표준자료 변환

기후모델 자료를 생산하는 모델링 그룹의 산출물은 자신의 모델에 지정된 저장 규칙에 따라 개별적인 방식으로 저장되어 진다. 하나의 파일에는 한 개 이상의 변수가 포함되어 있을 수 있으며, 저장 디렉토리 구조나 명명 규칙 또한 모델링 그룹에 의해 지정된 방법을 사용한다. 표준화되지 않은 방법으로 저장된 자료는 다양한 분야의 여러 연구자들이 공유하기에는 많은 문제점이 존재한다. 이에 CMIP에 자료 제출을 위해서는 MIP에서 요구되는 파일 구조와 메타데이터가 기후 커뮤니티의 표준 모델 실험 요건을 충족해야 한다. 즉, 개별 모델링 그룹은 데이터를 서로 다른 방식으로 저장하지만 CMOR(Climat Model Output Rewriter)(<https://cmor.llnl.gov>)를 사용

하여 데이터를 MIP에서 요구하는 형식으로 변환하여야 한다.

CMOR는 다양한 MIP에 데이터를 제공하는 모델링 센터의 노력을 줄이기 위해 설계되었다. 이것은 원본 데이터 구조에서 수행해야 하는 변환을 최소화하여 MIP 요구사항 충족시키는 것을 목적으로 한다. 기후 데이터를 교환하기 위해 CMOR 라이브러리를 적용하면 MIP 참여가 용이하다. 하나의 MIP에 출력 요건을 충족시키는 방법을 적용 시킨다면, 다른 MIP에 대한 출력을 준비하기가 쉬워진다. 국립기상과학원에서는 포트란과 셸 스크립트, 파이썬을 이용한 후처리과정을 개발하여 국제표준자료 변환 과정을 간편화하였다. CMOR 라이브러리를 이용하여 개발된 후처리 과정은 CMIP에서 정한 규칙에 따라 제출 변수를 NetCDF 형식의 국제 표준 자료로 손쉽게 변환할 수 있도록 해준다. CMOR 라이브러리는 C언어, 파이썬, 포트란 등 다양한 언어로 제공되고 있으며, 각 모델링 센터에서 자유롭게 원하는 라이브러리를 선택하여 사용할 수 있다. JSON 형태로 제공되는 ‘사용자 입력 파일’을 입력 자료로 사용하여 제출 변수에 필요한 필수 메타정보를 입력한다. 이 파일은 사용자 제공 메타데이터 및 구성 지시어를 제공하고 있다. 그 외 CMOR에 의해 작성될 수 있는 각 변수에 대해 필요한 메타 데이터의 대부분을 제공하는 ‘CMOR 테이블’을 이용하여 표준 출력 자료를 생산한다. 이 파일은 CMIP 공식 사이트(<https://cmor.llnl.gov>)에서 제공되며 사용자에게 의해 임의로 수정될 수 없다. CMIP에 참여하는 모델링 그룹에서 생산한 기후 모델 자료는 ESGF 분산 데이터저장소에 저장된다. 여기에 저장된 시나리오 자료는 다양한 분야에서 기후 과학을 연구하는 연구자를 위해 공개된다. 이에 각 기관에서 제출하는 자료는 국제표준자료 포맷의 기준에 적합해야 하는 것뿐만 아니라 자료의 품질 보증 과정을 통한 검증이 필요하다.

3.2.2 품질 보증 과정 수행

자료가 생산된 후에도 모델 출력에서 모델의 구성, 생산 도중에 이루어지는 절차와 선택까지 추적할 수 있는 있어야 한다. CMIP 커뮤니티 그룹에서는 자료 품질에 대한 표준 메트릭스를 개발하려는 노력이 이어졌다(Eyring et al., 2016b; Gleckler et al., 2016). 데이터 사용을 기록하고 추적할 필요성과 함께 과학적 재현성과 책임에 대한 필요성이 요구된다. 품질 보증(Quality Assurance, 이하 QA)은 국제 표준 규격 및 각 프로젝트 규칙을 준수하고 있음을 검사하는 과정으로 자료에 대한 신뢰성을 확보할 수 있게 해준다. 이에 국제표준자료로의 변환과정에서 CMIP6에서 제공하는 CMIP6 CVs (Controlled Vocabularies, 이후 CV)에 기반 한 메타데이터를 포함하고 있는지 확인한다. QA는 CMIP6 데이터 요구사항

명세를 잘 따르고 있는지, CV와 명명 규칙을 준수하는지, 공통 디렉토리 구조를 따라서 구성되어 있는지를 확인하는 과정이다. QA는 데이터 수명 주기 전체에 영향을 준다. 모든 단계에서 과학적 재현성을 가능하게 하는 출처 정보가 포함되어있는 가를 확인한다(Juckes et al., 2015). 데이터의 오류를 보고하지 않은 채로 방치할 경우 지구 시스템 모델링 커뮤니티 외부에 대한 데이터 신뢰도를 훼손할 수 있으므로 이러한 문제를 조기에 발견하는 것은 중요하다. 데이터 및 메타데이터가 모델의 시뮬레이션을 정확하게 반영하여 과학적 목적을 위해 안정적으로 사용될 수 있도록 해야 한다. QA에서 중요한 요소는 분산된 데이터에 공통 CV에 기초한 공통 메타데이터가 포함됨을 확인하는 것이며, 이는 서로 다른 그룹과 기관의 데이터에 대한 일관된 처리를 가능하게 할 것이다.

QA 절차는 기본적으로 데이터 생산자가 책임을 진다. 이에 많은 모델링 센터들은 CMIP 제출을 위한 자료 출력을 처리할 때 CMIP6 요구사항을 준수 하도록 이러한 단계를 워크플로우에 직접 포함되도록 하고 있다. 국립기상과학원에서도 이러한 QA과정을 포함시켜 후처리과정을 구축하였다. 사용한 도구는 독일의 기후컴퓨팅 센터에서 개발한 유틸리티인 QA-DKRZ (Quality Assurance-Deutsches Klima Rechen Zentrum, <https://qa-dkrz.readthedocs.io/en/latest/introduction.html>)이며, 변환자료의 메타정보를 검증하는 역할을 한다. QA-DKRZ는 파이썬으로 작성된 스크립트 기반 도구로 자료에 포함된 메타정보, 데이터 정보를 확인하여 오류 여부를 로그파일 형태로 제공한다.

QA는 CMIP6 품질보증 절차에 따라 ESGF 시스템 내에서도 수행 된다. 데이터를 인증하는 추가 QA 검사를 수행하여 인용 및 장기 보관에 적합하도록 한다. 데이터는 배포된 후에 오류가 발견될 수도 있다. 이에 ESGF에서는 Errata 서비스를 제공한다. 이미 배포된 자료에서 발견된 오류의 심각성 정도 (Critical, High, Medium, Low)와 함께 오류 정보를 ES-DOC에 올려서 자료를 이용하는 사용자들과 공유하고 있다. Errata는 치명적인 오류가 발생한 자료에 대해서는 공개를 철회하고 수정된 자료로 배포하도록 하는 서비스이다.

3.3 CMIP6 자료 제출을 위한 ESGF 노드 구축

3.3.1 ESGF 아키텍처

ESGF는 분산 통합 소프트웨어 아키텍처를 기반으로 한다. 이 시스템은 전 세계에 지리적으로 떨어져 있는 다수의 노드(Table 2)로 구성되지만 공통의 서비스, 프로토콜, API를 채택하여 상호 운용할 수 있도록 하였다. 데이터와 메타데이터는 각 노드에서 독립적으로 관리 및 저장되고 있으나

클라이언트는 하나의 저장소에 자료가 저장된 것처럼 접근 가능하다(Cinquini et al., 2014). ESGF는 CMIP5를 위한 IT 인프라를 제공하는 것을 목적으로 시작되었다. ESGF 소프트웨어 컴포넌트는 서로 다른 기관과 프로그램에 의해 지원을 받는 전 세계 많은 연구소가 공동 협력하여 개발하였다. ESGF를 구성하는 컴포넌트 모듈은 서로 상호 작용하고, 잘 정의된 인터페이스를 통해 클라이언트와도 상호 작용하도

록 만들어졌다. 또한 구현 언어에 관계없이 다른 컴포넌트가 시스템에 쉽게 포함될 수 있도록 하였다. ESGF 노드는 데이터 및 메타데이터 액세스와 사용자 관리를 가능하게 하는 서비스와 애플리케이션으로 구성된다. ESGF 서비스는 데이터 노드, 인덱스 노드, 인증 제공 서비스, 계산 노드의 4가지 기능의 논리적 영역으로 분류되어 있다(Fig. 3).

Table 2. ESGF data nodes status. (Data from <https://esgf-node.llnl.gov/search/cmip6>, accessed 2020 May 30)

| No | Data Node | No | Data Node |
|----|--------------------------------|----|------------------------------|
| 1 | 145.100.59.180.surf-hosted.nl | 26 | esgf-data1.llnl.gov |
| 2 | acdisc.gesdisc.eosdis.nasa.gov | 27 | esgf-data2.ceda.ac.uk |
| 3 | aims3.llnl.gov | 28 | esgf-data2.diasjp.net |
| 4 | cmip.bcc.ema.cn | 29 | esgf-data2.llnl.gov |
| 5 | cmip.dess.tsinghua.edu.cn | 30 | esgf-data3.ceda.ac.uk |
| 6 | cmip.fio.org.cn | 31 | esgf-data3.diasjp.net |
| 7 | cordexesg.dmi.dk | 32 | esgf-dev.bsc.es |
| 8 | crd-esgf-drc.ec.gc.ca | 33 | esgf-ictp.hpc.cineca.it |
| 9 | data.meteo.unican.es | 34 | esgf-nimscmp6.apcc21.org* |
| 10 | dataserver.nccs.nasa.gov | 35 | esgf-node.cmcc.it |
| 11 | dist.nmlab.snu.ac.kr | 36 | esgf-node2.cmcc.it |
| 12 | dpegf03.nccs.nasa.gov | 37 | esgf.anl.gov |
| 13 | esg-cccr.tropmet.res.in | 38 | esgf.apcc21.org |
| 14 | esg-dn1.nsc.liu.se | 39 | esgf.bsc.es |
| 15 | esg-dn1.ru.ac.th | 40 | esgf.dwd.de |
| 16 | esg-dn2.nsc.liu.se | 41 | esgf.ichec.ie |
| 17 | esg.camsma.cn | 42 | esgf.nci.org.au |
| 18 | esg.lasg.ac.cn | 43 | esgf.rcec.sinica.edu.tw |
| 19 | esg.pik-potsdam.de | 44 | esgf1.dkrz.de |
| 20 | esg1.umr-cnrm.fr | 45 | esgf2.dkrz.de |
| 21 | esgdata.gfdl.noaa.gov | 46 | esgf3.dkrz.de |
| 22 | esgf-cnr.hpc.cineca.it | 47 | gpm1.gesdisc.eosdis.nasa.gov |
| 23 | esgf-data.ucar.edu | 48 | noresg.nird.sigma2.no |
| 24 | esgf-data1.ceda.ac.uk | 49 | vesg.ipsl.upmc.fr |
| 25 | esgf-data1.diasjp.net | | |

Note: ESGF nodes built by NIMS-KMA

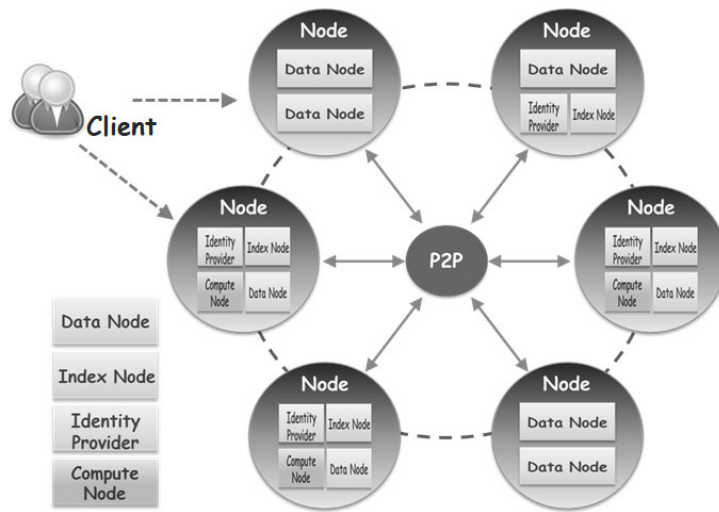


Fig. 3. ESGF architecture. The ESGF employs a system of distributed peer nodes using P2P protocol. (Fig. 1 in Cinquini et al. (2014). Refer to corresponding caption for further details)

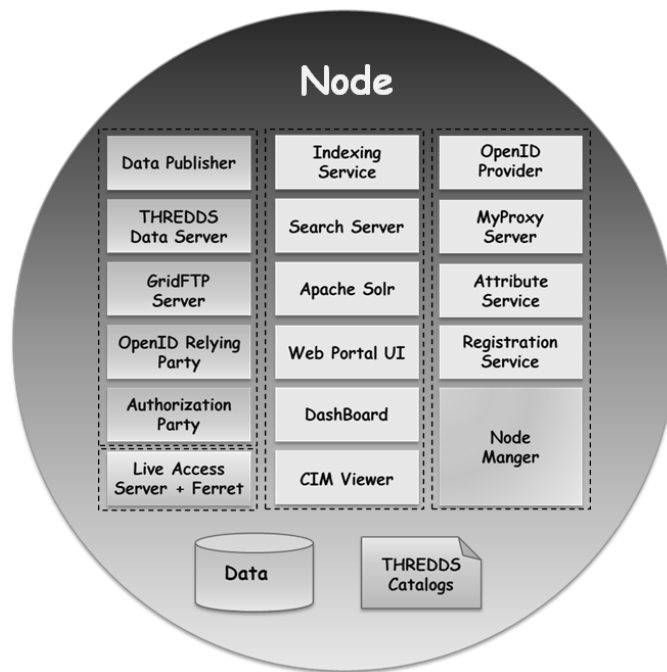


Fig. 4. ESGF Node software stack. (Fig. 2 in Cinquini et al. (2014). Refer to corresponding caption for further details)

ESGF에서 제공하는 서비스는 Table 3에서 보여준다. 데이터 노드는 안전한 데이터 게시 및 액세스를 위한 서비스 제공을 담당한다. 인덱스 노드는 메타데이터의 인덱싱 및 검색을 위한 서비스를 포함하고 있다. 인증 제공서비스는 사용자 인증 및 사용자 속성의 안전한 전달이 가능하도록 해준다. 계산 노드는 데이터 분석 및 시각화를 위한 높은 수준의 서비스를 제공한다.

Table 3. The ESGF services. They are divided into four areas of functionality.

| Service | Description |
|-------------------|--|
| Data Node | Secure data publication and access |
| Index Node | Indexing and searching metadata, implemented using Apache Solr |
| Identity Provider | User authentication and secure delivery of user attributes |
| Compute Node | Higher-level services for data analysis and visualization |

소프트웨어 스택에는 ESGF가 직접 개발한 구성 요소와 Postgres, Tomcat(Vukotic and Goodwill, 2011), Solr(Smilely et al., 2015)과 같은 보편적으로 사용되는 오픈 소스나 타사 응용프로그램, MyProxy, GridFTP와 같은 eScience 프로그램 및 TREDDs(John Caron and Davis, 2006), 라이브 액세스 서버(Live Access Server, LAS)(Hankin et al., 2001) 등이 포함되어 있다. (Fig. 4)

ESGF 소프트웨어 스택은 노드 관리자, P2P프로토콜, 데이터 작성 및 공개 어플리케이션, 검색 서비스, 보안 인프라, 웹 포털 사용자 인터페이스, 데이터 액세스, 데이터 분석 및 시각화, 대시보드, CMI 뷰어를 주요 구성 요소로 포함하고 있다.

ESGF 시스템 구조는 여러 프로토콜과 서비스를 통하여 사용자들이 데이터를 다운로드 할 수 있도록 해준다. 데이터에 접근 가능하도록 하는 메커니즘은 데이터가 저장된 호스트의 노드 관리자에 의해 설치된다. ESGF에 저장된 데이터는 인터넷 브라우저를 통한 접근, OPenDAP, Wget, GridFTP, Globus Online과 같은 방법으로 자료에 접근하여 이용할 수 있다. 시스템 관리자는 각 노드에서 서비스에 맞는 설정 정보 파일을 작성 한다.우리는 노드 구축을 위해 이러한 설정 파일의 시스템 구성 정보를 이용하여 서비스를 설정하였다. ESGF의 자료 제공을 위한 프로토콜과 서비스는 다음과 같다.

- 하이퍼링크를 이용하여 사용자는 브라우저에서 직접적

으로 다운로드 할 수 있다.

- Open Data Access Protocol(OPeNDAP)(Cornillon et al., 2003)은 사용자가 변수, 시간적, 지리적 범위에 의해 맞춤형 데이터 선택이 가능하도록 하는 개념이 적용되었다. 이 서비스는 일반적으로 파일의 크기가 2GB 이하로 제한되어 있다.
- Wget 스크립트를 이용하는 사용자는 파일이 저장된 위치나 이름을 알지 못하더라도 접근 가능 하다. 또한 사용자가 스크립트를 수정하여 사용 가능한 장점이 있다.
- GridFTP 액세스(Allcock et al., 2005)는 FTP에 기반 한 특수 프로토콜을 사용한다. 이 서비스는 성능이 우수하고 인증이 확실하며, 신뢰할 수 있는 데이터 전송을 허용한다.
- Globus Online은 GridFTP를 통해 안정적으로 파일 전송 기능을 제공하는 호스트 서비스다.

ESGF는 주로 데이터 저장 및 접근을 위한 글로벌 인프라를 제공하는 데 초점을 맞추고 있다. 그러나 기후 데이터의 규모와 범위가 계속 확대됨에 따라 데이터가 위치한 서버에서 수행되는 데이터 분석의 필요성이 점점 뚜렷해지고 있다. 이에 따라 Live Access 서버 엔진, UV-CDAT, RCMET 등의 툴을 사용하여 서버 및 클라이언트 측 모두에서 사용할 수 있는 데이터 분석 및 시각화 기능을 제공하고 있다.

3.3.2 자료 제출을 위한 노드구축

ESGF에 자료를 제출하기 위해 각 모델링 센터는 이미 구축되어 있는 노드에 자료를 전송할 수도 있고, 자신의 노드를 구축하여 자료를 저장 할 수도 있다. 자료의 용량이나 기존 ESGF 노드로의 자료 전송 속도 등을 고려하여 선택할 수 있다. K-ACE를 이용하여 산출된 자료를 저장 및 관리하기 위해 우리는 자체 노드를 구축하였다. 이 노드(<https://esgf-nimscmp6.apcc21.org>)(Fig. 5)는 APEC 기후센터(APEC Climate Center, 이후 APCC) 협조를 받아 APCC 네트워크망을 이용하고 있다. 자료 서비스를 위하여 제출할 자료의 용량 및 관리의 용이성 등을 고려하여 구축된 국립기상과학원 자체 노드는 Fig. 6과 같이 구성되어 있다.

ESGF 노드 구축을 위해서는 필요한 시스템 및 네트워크 요구사항을 만족해야 한다. 서버는 제어 서버와 관리 서버 2대가 요구된다. 그러나 반드시 물리적으로 독립된 서버일 필요는 없다. ESGF 노드는 리눅스 계열 CentOS를 기본 운영체제로 이용하고 있으며, 오픈 소스 기반 자동화 관리 도구인 ESGF-Ansible을 사용 한다. Table 4는 우리가 구축한 노드 서버의 시스템 환경 및 설치에 요구되는 필수 프로그램

램 목록이다.

데이터의 제출 및 관리에 사용되는 기본 시스템 및 유틸리티는 파이썬 라이브러리를 사용하여 작성되어 있으므로 파이썬 2.7 이후 버전이 필요하며, HTTPS를 통한 웹서비스가 가능하도록 보안 인증서(SSL Certificates)가 설치되어 있어야 한다. ESGF 노드의 설치가 완료되면 ESGF에 CMIP6 자료 제출이 가능하다. 제출 과정은 Fig. 6에서 보여준다. 기후 모델에서 산출된 자료는 후처리를 통해 앞 절에서 설명한 국제표준자료 변환 과정과 QA과정을 거친 후 로컬 서버의 스토리지에 저장한다. 제출을 위해서는 mapfile을 생성하여야 한다. CMIP 제출 변수 파일에 각각에 관련 mapfile이

생성되는데, 이 파일에는 저장 위치, 오류 체크를 통해 무결성을 확인하기 위한 체크섬, 파일의 버전 정보 등을 담고 있다. 이렇게 생성된 mapfile들을 이용하여 PostgreSQL DB, Thredds catalog, Index node에 순차적으로 정보를 저장시켜 최종적으로 ESGF 노드에서 자료를 활용할 수 있도록 제출 프로세스가 수행되어진다. 이러한 과정을 통해서 K-ACE모델의 CMIP6 변수 190여 종, 약 3,200개를 ESGF 노드에 제출하여 공개하고 있다. 이 외에 UKESM1모델과 CORDEX 자료도 ESGF 노드에 제출하여 미래 기후변화 예측 연구에 도움을 주고자 한다.

Table 4. System Specification

| Type | Properties |
|-----------|--|
| Processor | Dell PowerEdge720 |
| CPU | Intel Xeon E5 |
| RAM | 64G |
| OS | CentOS Linux 7 |
| Program | Python 2.7 or later, ESGF Ansible(https://github.com/ESGF/esgf-ansible), ESGF v2.5 or later(https://github.com/ESGF/esg-publisher), commercial CA |

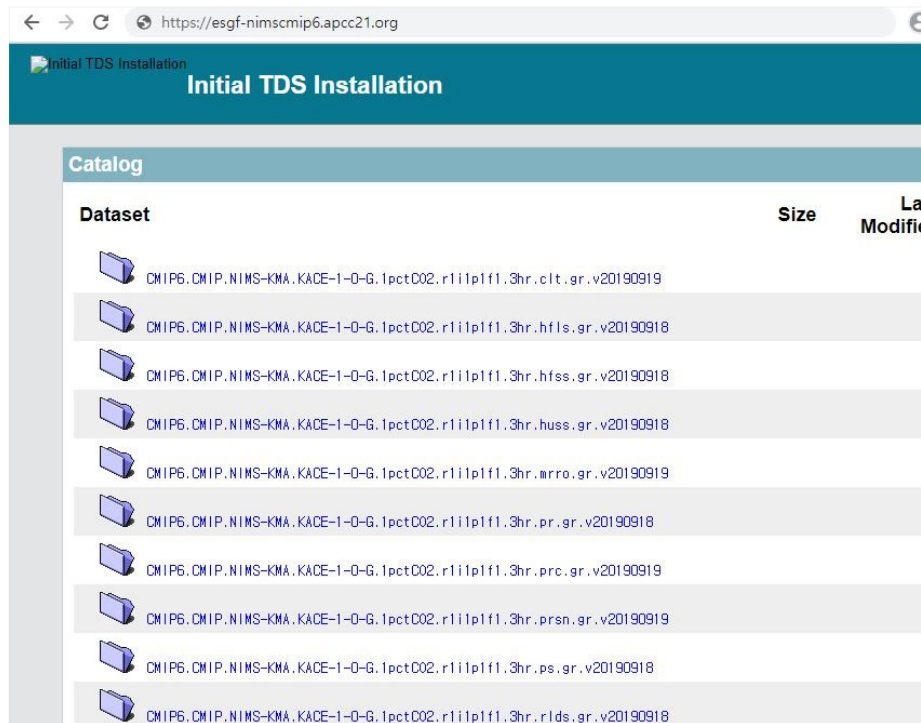


Fig. 5. Thredds(TDS) catalogs of NIMS-KMA node.
(<https://esgf-nimscmp6.apcc21.org>, accessed 2020 May 30)

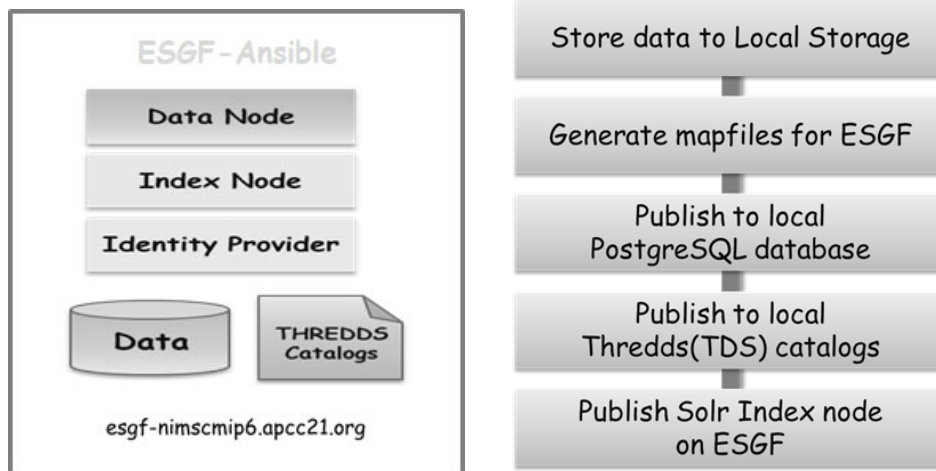


Fig. 6. NIMS-KMA node architecture and overview of the ESGF publication process. NIMS-KMA node built using ESGF-Ansible, esgf-publisher(left), Steps to publish climate model output on ESGF(right)

4. 요약 및 결론

국립기상과학원에서는 2007년부터 CMIP에 참여하여 4차, 5차 평가보고서에 시나리오 산출을 통해 기여하여 왔다. 2022년 발간 예정인 IPCC 6차 평가보고서에도 참여하기 위해 CMIP6에 자료를 생산 및 제출함과 더불어, 국제데이터 노드를 구축하고, ESGF 노드로서의 역할을 수행 중이다.

ESGF에 제출된 자료는 기후 변화를 연구하는 많은 연구자들이 이용하여 과학적 질문에 대한 해답을 제공하는 기본 자료로 사용되고 있다. 기후 모델을 생산하는 모델링 센터에서의 산출물은 자신의 모델에 지정된 저장 규칙에 의해 각자의 방식으로 자료를 저장되고 관리되어진다. 자료 사용자 측면에서 이것은 자료 이용에 많은 불편함을 초래한다. 이에 ESGF에 CMIP자료 제출을 위해서는 국제표준 자료형식인 NetCDF로의 변환 및 자료에 대한 정보를 담고 있는 메타데이터가 포함되도록 요구하고 있다. 이에 국립기상과학원에서는 국제표준 자료변환 과정(CMORization) 수행하였다. 또한 자료에 대한 과학적 재현성과 기초적인 내구성, 데이터의 출처에 대한 입증을 보증하기 위해 수행되어진 품질 보증 과정(QA)과정은 생산된 변수 자료를 검증하여 신뢰성을 높였다.

전 세계 많은 나라에서 생산된 CMIP자료는 ESGF 분산 데이터 노드에 저장되어 기후과학자 및 관련 연구자들에게 공유되어진다. P2P방식으로 공유되고 있는 CMIP 자료는 여러 나라의 노드에서 관리되고 있다. CMIP에 참여하는 모

델링 센터에서는 자료 제출을 위해 이미 구축되어 있는 ESGF 노드를 이용할 수도 있고 자체 노드를 구축할 수 있다. 국립기상과학원에서는 WIP의 권고안을 준수한 국제 데이터노드(<https://esgf-nimscmip6.apcc21.org>)를 구축하여 CMIP6 자료를 제출하였다. 노드는 ESGF 포탈(<https://esgf-node.llnl.gov>)에 연결되어 CMIP6 자료 제공에 이용되고 있다. IPCC 5차 평가보고서에 비해 6차 평가보고서를 위해 산출되는 자료 용량은 크게 증가하였다. Cinquini et al.(2014)에 의하면 이전에 비하여 훨씬 높은 해상도의 보다 정교한 기후 모델 자료와 관측 데이터를 포함하고 있는 CMIP6는 350PB에서 3EB(Exabyte) 사이의 데이터가 저장될 것으로 예상하였다. 이는 CMIP5에 비하면 최소 100여배 이상 크게 증가된 용량이다. 자료는 생산하는 만큼, 대용량으로 처리, 배포, 관리하는 자료 유통, 전달이 매우 중요하게 되었다. 이에 ESGF를 통한 산출에서 제공까지의 시간을 단축한 신속한 자료제공으로 국내외 사용자들의 자료 활용에 크게 기여할 수 있으리라 기대한다.

사사

이 연구는 기상청 국립기상과학원 「AR6 기후변화시나리오 개발·평가(KMA2018-00321)」의 지원으로 수행되었습니다.

References

- Allcock W, Bresnahan J, Kettimuthu R, Link M. 2005. The Globus striped GridFTP framework and server. SC05:Proceedings of the 2005 ACM/IEEE conference on Supercomputing; IEEE. p. 54.
- Balaji V, Taylor KE, Juckes M, Lawrence BN, Durack PJ, Lautenschlager M, Blanton C, Cinquini L, Denvil S, Elkington M. 2018. Requirements for a global data infrastructure in support of cmip6. *Geoscientific Model Development*. 11(9): 3659-3680.
- Cinquini L, Crichton D, Mattmann C, Hamey J, Shipman G, Wang F, Ananthkrishnan R, Miller N, Denvil S, Morgan M, Pobre Z, Bell GM, Doutriaux C, Drach R, Williams D, Kershaw P, Pascoe S, Gonzalez E, Fiore S, Schweitzer R. 2014. The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems*. 36:400-417.
- Collins WJ, Lamarque J-F, Schulz M, Boucher O, Eyring V, Hegglin MI, Maycock A, Myhre G, Prather M, Shindell D. 2017. Aerchemmp: Quantifying the effects of chemistry and aerosols in cmip6. *Geoscientific Model Development*, 10(2): 585-607.
- Cornillon P, Gallagher J, Sgouros T. 2003. OPeNDAP: Accessing data in a distributed, heterogeneous environment. *Data Science Journal*. 2: 164-174.
- Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE. 2016a. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*. 9(5): 1937-1958.
- Eyring V, Gleckler PJ, Heinze C, Stouffer RJ, Taylor KE, Balaji V, Guilyardi E, Joussaume S, Kindermann S, Lawrence BN et al. 2016b. Towards improved and more routine Earth system model evaluation in CMIP. *Earth System Dynamics*. 7(4): 813-830.
- Gates WL. 1992. AN AMS CONTINUING SERIES: GLOBAL CHANGE-AMIP: The Atmospheric Model Intercomparison Project. *Bulletin of the American Meteorological Society*. 73(12): 1962-1970.
- Gleckler PJ, Doutriaux C, Durack PJ, Taylor KE, Zhang Y, Williams DN, Mason E, Servonnat J. 2016. A more powerful reality test for climate models. *Eos*. 97.
- Hankin S, Callahan J, Sirott J. 2001. An Overview of the Live Access Server. AGUSM. 2001:A31D-10.
- IPCC. 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change: Cambridge University Press.
- John C, Davis E. 2006. UNIDATA's THREDDS data server. 22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology
- Juckes M, Eyring V, Taylor K, Balaji V, Stouffer R. 2015. The CMIP6 Data Request: The next generation climate archive. EGUGA. 13112.
- Lee J, Kim J, Sun M-A, Kim B-H, Moon H, Sung HM, Kim J, Byun Y-H. 2020. Evaluation of the Korea Meteorological Administration Advanced Community Earth-System model (K-ACE). *Asia-Pacific Journal of Atmospheric Sciences*. 56(3): 381-395.
- O'Neill BC, Kriegler E, Riahi K, Ebi KL, Hallegatte S, Carter TR, Mathur R, van Vuuren DP. 2014. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*. 122(3): 387-400.
- O'Neill BC, Tebaldi C, van Vuuren DP, Eyring V, Friedlingstein P, Hurtt G, Knutti R, Kriegler E, Lamarque J-F, Lowe J et al. 2016. The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*. 9(9): 3461-3482.
- Sellar AA, Jones CG, Mulcahy JP, Tang Y, Yool A, Wiltshire A, O'Connor FM, Stringer M, Hill R, Palmieri J et al. 2019. Ukesml: Description and evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*. 11(12): 4513-4558.
- Smiley D, Pugh E, Parisa K, Mitchell M. 2015. Apache solr enterprise search server. Packt Publishing Ltd.
- Vukotic A, Goodwill J. 2011. Apache tomcat 7. Springer.
- Williams DN, Lawrence BN, Lautenschlager M, Middleton D, Balaji V. 2011a. The earth system grid federation: Delivering globally accessible petascale data for CMIP5. *Proceedings of the Asia-Pacific Advanced Network*. 32(0).
- Williams DN, Taylor KE, Cinquini L, Evans B, Kawaniya M, Lautenschlager M, Lawrence B, Middleton D, Contributors E. 2011b. The Earth System Grid Federation: Software framework supporting CMIP5 data analysis and dissemination. 56(2): 40-42.