

Kendini tanımlayan yer sistem modellerine doğru: Bir iş akışı uygulaması

Ufuk Utku TURUNÇOĞLU*, Hasan Nüzhet DALFES

İTÜ Bilişim Enstitüsü, Hesaplamalı Bilim ve Mühendislik Programı, 34469, Ayazağa, İstanbul

Özet

Bilimsel iş akış sistemlerinin karmaşık bileşen ve süreçler içeren çalışmalarda kullanılması, günümüzde oldukça popüler bir araştırma alanıdır. Özellikle yer sistem bilimleri açısından bakıldığında, birbirlerine sıkı bir şekilde bağlı, birçok alt bileşen içeren yer sistem modellerine ait süreçlerin basitleştirilmesi ve modüler bir yapıya kavuşturularak standartlaştırması için, bilimsel iş akış sistemleri kullanılabilir. Bu çalışmanın asıl amaçlarından biri, karmaşık yer sistem modelleme uygulamalarının, yüksek başarımlı hesaplama sistemlerinde çalıştırılmasının gittikçe zorlaşması, kendi kendisini tanımlayan yer sistem modellerine ve varolan karmaşık modelleme sistemlerinin anlamlı bir şekilde soyutlaştırılarak ortalama bir kullanıcı seviyesine indirgenmesine duyulan ihtiyaçtır. Bu amaçla, literatürde sıkça kullanılan bilimsel iş akışı ve çerçeve yaklaşımlarının birleştirilerek standartlaştırılmış bir çalışma ortamının yaratılması için yeni bir metodoloji geliştirilmiştir. Önerilen metodolojinin sınanması amacıyla, gerçekçi bir yer sistem modelleme uygulaması kullanılmış ve sonuçları analiz edilmiştir. Bu amaçla geliştirilen uygulama örneği, çerçeve yaklaşımı kullanılarak birleştirilmiş bir atmosfer-okyanus model sisteminin, bölgesel iklim sistemini daha gerçekçi olarak modellemek amacıyla iş akış sistemine entegre edilmesi ve model bileşenlerinden köken bilgisinin otomatik olarak toplanması süreçlerini içermektedir. Sonuçlar iş akışı ortamına entegre edilen model sistemlerinin daha kolay kullanılabilirliğini ve toplanan köken bilgisinin çalışmanın evrimini kayıt altına alması açısından son derece önemli olduğunu göstermektedir. Ayrıca modellerin kuple edilmesi için kullanılan çerçeve uygulaması (ESMF), her bir alt yer sistem modeli için standart bir arayüz tanımlayarak model kuple sürecini ve yönetimini kolaylaştırmaktadır.

Anahtar Kelimeler: Bilimsel iş akışı, çerçeve yaklaşımı, köken bilgisi, Kepler, ESMF.

*Yazışmaların yapılacağı yazar: Ufuk Utku TURUNÇOĞLU. u.utku.turuncoglu@be.itu.edu.tr; Tel: (212) 285 71 94 (dahili: 142). Bu makale, birinci yazar tarafından İTÜ Bilişim Enstitüsü, Hesaplamalı Bilim ve Mühendislik Programında tamamlanmış olan "Yer sistem modellemesi için iş akışı tasarımı ve yönetim sistemi" adlı doktora tezinden hazırlanmıştır. Makale metni 26.05.2010 tarihinde dergiye ulaştırılmış, 10.06.2010 tarihinde basım kararı alınmıştır. Makale ile ilgili tartışmalar 31.08.2011 tarihine kadar dergiye gönderilmelidir. Bu makaleye "Turunçoğlu, U U., Dalfes, H S., (2011) 'Uygurcadan Türkçeye bilgisayarlı çeviri', İTÜ Dergisi/D Mühendislik, 10: 3, 105-115" şeklinde atıf yapabilirsiniz.

Towards self-describing earth system models: a workflow application

Extended abstract

The application of scientific workflow systems for orchestrating complex tasks is still an open research area. In particular, earth system related modeling applications consists of different tasks that are closely related to each other and scientific workflow systems can be used to simplify these sub processes and their relationships. The motivation for this work is driven by the complexities of running a large modeling system on a high performance computing and network systems and need to reduce those complexities, particularly for the average user.

In this study, it is presented and analyzed a new methodology to combine scientific workflow system and modeling framework approach together to create standardized work environment. Then, the proposed methodology is tested using a typical and realistic earth system modeling application. The result of example workflows that are based on the proposed methodology is a part of this study.

A Modeling framework is a standardized programming environment for combining model components and couplers of different kinds of earth system models using a common calling interface. Earth System Modeling Framework (ESMF) is one of the most popular examples for this approach. It consists of a superstructure for coupling components of Earth system applications and an infrastructure of robust, high-performance utilities and data structures that ensures consistent component behavior. Extending beyond the modeling framework approach, scientific workflow systems create standardized interfaces to a variety of technologies and automate the execution and monitoring of a heterogeneous workflow. Namely, a scientific workflow system is a problem-solving environment that simplifies tasks by creating meaningful and easy understandable sub-tasks and combining them to form executable data management and analysis pipelines.

Kepler was chosen as the workflow environment for this work because it is open source, platform independent, and it supports different models of computation such as Process Network (PN) and Synchronous Data Flow (SDF). Kepler is a scientific workflow system that is based on the Ptolemy II project. The actor-oriented design of Kepler enables users to

create hierarchically structured scientific workflows. The overall execution of model is controlled by separate component that is called a director (a special actor). Kepler supports different model of computation types via its directors.

In application example, a regional coupled climate modeling system is developed for Mediterranean region and integrated into workflow system to provide better representation of regional climate system. This application has crucial importance in downscaling output of the global circulation models over Turkey and near regions. It also can be used to create better representation of regional climate for the future scenarios. The workflow application also collects provenance information automatically from the coupled earth system modeling system to reproduce, compare and debug results. The coupled atmosphere-ocean modeling system is based on NCAR's Weather Research and Forecasting (WRF) and Regional Ocean Modeling System (ROMS) model.

To couple atmosphere and ocean models the ESMF library is used. For the atmospheric component (WRF), the experimental ESMF-IO structure is used to create ESMF fields and states to share data between different model components. The WRF ESMF-IO is also modified to add field level metadata into coupling variables (heat and momentum fluxes) of atmospheric model. In ocean component (ROMS), the coupler code is written to control each model component and also the gridded component code is developed to run each earth system model via ESMF init, run and finalize methods. The gridded component level metadata is added into coupler to create prototype version of self-describing modeling system.

The results show that the developed workflow environment is capable of running different earth system models on a different high performance computing resource with a meaningful abstraction. The proposed work environment acts as an abstraction layer and hides the detail of the used infrastructure and earth system model from user and it also collect standardized provenance information about both model and computing resource to represent the work environment as possible as it can.

Keywords: Scientific workflow, model coupling, provenance information, Kepler, ESMF.

Giriş

Bilgi teknolojileri, yüksek başarımlı hesaplama ve depolama sistemleri, yeni ölçüm ağlarının geliştirilmesi ve disiplinler arası uygulama alanlarının yaygınlaşmasında yaşanan son yıllardaki hızlı ve düzenli gelişim ile birlikte, yer sisteminin daha iyi anlaşılmasına yönelik analiz ve modelleme çalışmaları, gün geçtikçe daha karmaşık ve hesaplama yönünden yoğun problemlerin çözümüne odaklanmaktadır.

Çözülme istenen problemin karmaşıklığı ve birçok alt bileşen ve süreç içermesi araştırmacıları birbirinden farklı düzeyde bilgi ve tecrübe gerektiren birçok farklı türde matematiksel ve fiziksel modeli, hesaplama ortamını ve uygulamayı birlikte kullanmaya itmektedir. Ancak bu durum araştırmacının probleme odaklanmasını engellemekte ve çalışmanın asıl amacı olan problemin çözümü ve analizi sürecini yavaşlatmakta veya imkansız hale getirmektedir.

Yukarıda tanımlanan problemlerin çözümüne yönelik birçok farklı yaklaşım bulunmaktadır. Bu problemler karşısında üretilen çözüm önerilerini gruplamak istediğimizde “çerçeve” ve “iş akışı” yaklaşımı olarak isimlendirilen iki farklı yaklaşım olduğu görülmektedir.

Çerçeve yaklaşımı, yer sistem modellerinin birbirleri ile uyum içinde çalışmasını sağlamak amacıyla modellerin birbirleri ile olan arayüzlerini standart bir yapıya kavuşturmayı amaçlar. Bu yaklaşım için verilebilecek örnekler arasında MCEL (Bettencourt 2002), MCT (Larson vd., 2005 ve Craig vd., 2005) ve ESMF (Hill vd., 2004 ve Collins vd., 2005) sayılabilir. ESMF (Earth System Modeling Framework) kütüphanesi uygulamasına örnek olarak CCSM (Community Climate System Model) modelinin MCT uygulaması ile yazılmış kuple bileşeninin ESMF ile tekrar geliştirilmesi verilebilir. Bu şekilde küresel modelin daha performanslı bir şekilde çalışması ve diğer bölgesel modeller ile kolayca kuple edilmesi amaçlanmaktadır.

Çerçeve yaklaşımının aksine bilimsel iş akışı uygulamaları daha önce değinilen problemleri çözmek için, tüm süreci birbirleri ile ilişkili ve

daha kolay anlaşılır birçok anlamlı parçaya ayırıp tanımlayan standart bir yönetim sistemi geliştirmeye odaklanmıştır (Bowers vd., 2005). Bu yöntem ile yer sistem model çalışmalarında sıklıkla karşılaşılan veri ön/son işleme, kullanılan modeli farklı hesaplama sistemlerine aktarma ve görselleştirme gibi farklı süreçlerin tek bir yapı altında birleştirilmesi amaçlanmaktadır. Bu tip bilimsel iş akışı geliştirme uygulamalarına Kepler (Ludäscher vd., 2006), Taverna (Oinn vd., 2004), Triana (Majithia vd., 2004) uygulamaları ile LEAD (Plale vd., 2006) projesi örnek verilebilir.

LEAD projesi yer sistem bilimleri açısından bakıldığında iş akışı yaklaşımını bilimsel ağ geçidi kavramı ile birleştirilmesi bakımından oldukça önemli bir örnektir. Bu bilimsel ağ geçidi uygulaması ile meteorolojik veriye ulaşma, analiz etme, modelleme ve görselleştirme aşamalarında kullanılan ileri seviyedeki birçok farklı araç tek bir merkez altından yönetilebilmektedir. İçerdiği iş akışı uygulaması ile WRF bölgesel atmosfer modelini TeraGrid hesaplama gridi üzerinde farklı veri setleri ile çalıştırıp model sonuçlarını incelemek mümkündür. Sistem ile belirli atmosferik şartlar sağlandığında ilgili bölge için otomatik olarak WRF modelinin çalışmasını sağlanabilmektedir ve bu sistem tayfunların izlenmesinde deneysel olarak kullanılmaktadır.

Kullanılan model sisteminin iş akışı uygulaması ile entegre edilmesi ve yer sistem modellerinin kendi kendisini tanımlayacak bir yapıya kavuşturulması amacıyla model ve içerdiği alt bileşenlerden otomatik olarak köken ve ön veri bilgisinin toplanması gereklidir. ESMF kütüphanesi içerdiği özellik nesnelere ile model ve model bileşenlerine ait ön veri bilgilerinin tanımlanmasını ve model çalıştığı anda bu bilgilerin belirli bir formda diske yazılmasını desteklemektedir. Bu şekilde modele ait alan değişkeni, model bileşeni ve model bileşenleri arasında değiştirilecek veri grupları bazında ön veri tanımlanması ve bu bilgilerin başka uygulamalar tarafından kullanılması mümkün olmaktadır.

Bu çalışmanın ileriki bölümlerinde geliştirilen metodoloji ve temel tanımlar hakkında bilgi ve

rilecek, sonraki bölümlerde ise metodolojinin sınanması amacıyla tasarlanan uygulama örneği üzerinden iş akışı ve çerçeve yaklaşımlarının birlikte kullanılarak kendini tanımlayan ve kup- le edilmiş bir yer sistem modelinin nasıl geliştirebileceği detaylı olarak okuyucuya sunulacaktır.

Temel tanımlar

Çerçeve ve iş akışı yöntemlerinin birleştirilerek yeni bir çalışma ortamının yaratılması ve bu ortamda kullanılan modellerin, kendi kendilerini tanımlayan bir şekilde çalışmalarını sağlamak amacıyla, köken ve ön veri bilgilerinin yer sistem modelinden otomatik olarak toplanması gerekliliği daha önceki bölümde okuyucuya aktarılmıştı. Bu bölüm geliştirilen yeni metodolojinin daha iyi anlaşılabilmesi için temel tanımları okuyucuya vermeyi amaçlamaktadır.

Köken bilgisi, basit olarak yapılan çalışmanın gelişimi ve evrimi ile ilgili tutulan kayıt olarak tanımlanabilir. İş akış sistemleri açısından incelendiğinde köken bilgisi, iş akış sistemleri tarafından yönetilen süreçlerin izlenmesi ve bunlar hakkında detaylı bilgi toplanmasını amaçlar. Köken bilgisi toplanan bilginin türüne göre dört ana gruba ayrılabilir (Bowers vd., 2006):

- Sistem köken bilgisi: Model benzetiminin çalıştığı uzak hesaplama sistemi hakkında bilgi içerir. Bu grup altında işletim sistemi, kullanılan uygulama ve kütüphanelerin sürümleri, tanımlanan çevre değişkenleri vb. hesaplama ortamına özgü tüm bilgiler sayılabilir. Bu şekilde sistem üzerinde yapılan değişiklikler kayıt altına alınmış olur.
- İş akışı uygulaması köken bilgisi: Bu başlık altında iş akış ortamının kendisi hakkında toplanan bilgiler bulunur. Tasarlanan iş akışının sürümü ve tasarım sürecinde geçirdiği evreler köken bilgisi olarak toplanır ve iş akışları arasındaki farklar ve buna bağlı hatalar kolayca takip edilebilir.
- Veri köken bilgisi: Temel olarak çalışma kapsamında kullanılan ve üretilen verinin oluşum süreci, yapısı, üzerinde yapılan işlemler ve kaynağı hakkında bilgi içerir. Bu

şekilde veri işleme sırasında oluşabilecek hatalar tespit edilebilir.

- Süreç köken bilgisi: İş akış sistemi tarafında çalıştırılan süreçler hakkında istatistiksel bilgiler içerir. Bu bilgiler arasında veri transfer hızı, transfer edilen veri miktarı gibi bilgiler bulunmaktadır. Bu bilgi iş akış sisteminin veya kullanılan yer sistem modelinin performansını arttırmak amacıyla kullanılabilir.

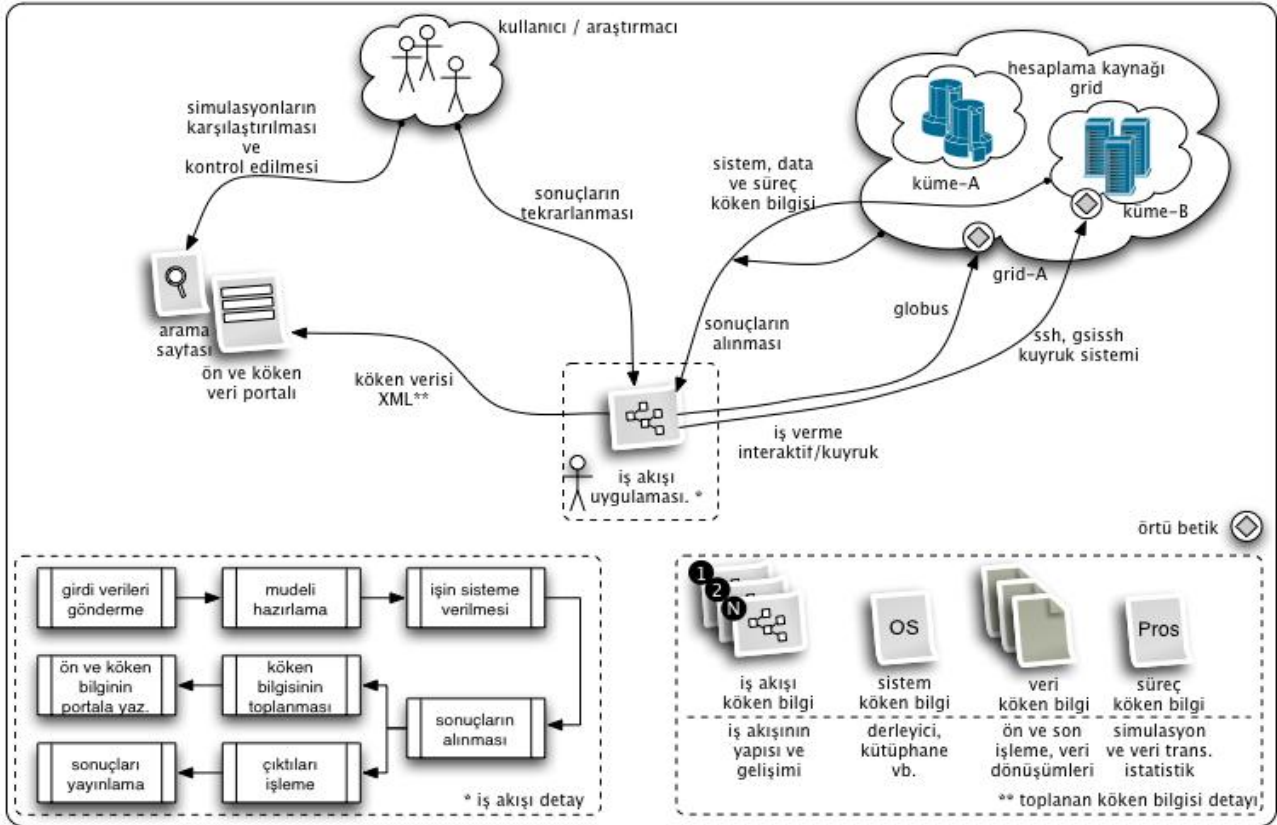
Yukarıda sıralanan her bir köken bilgisi tipi yapılan çalışmanın ve kullanılan ortamın tekrar üretilebilmesi açısından son derece önemlidir. Modelin derleme aşamasında köken bilgisinin toplanması amacıyla ORNL/NCSU tarafından geliştirilen bir dizi Python betiği, değiştirilerek sınama uygulamasından köken bilgisinin XML formunda toplanması için kullanılmıştır.

Ön veri, köken bilgisinden farklı olarak üretilen bilgiyi veya kullanılan model sistemini tanımlayan bir bilgi bütünüdür ve çoğunlukla veri hakkındaki veri olarak tanımlanır. Bu bilgi tipi yer sistem modellerinin kendi kendisini tanımlayan bir yapıda tasarlanabilmesi için kullanılabilir. Bu çalışma kapsamında, model sisteminden ön veri ve köken bilgisinin toplanması amacıyla ESMF uygulamasının özellik nesnelere kullanılmaktadır.

Yöntem

Yer sistem uygulamaları bir veya birden fazla modelin birlikte veya tekil olarak belirli bir problemin çözülmesi amacıyla kurulması, çalıştırılması ve sonuçların analiz edilmesi süreçlerini kapsamaktadır. Bunun gibi birçok farklı alt süreçten oluşan çalışma sistemlerinin iş akış uygulamaları ile kontrol edilmesi için bilşenler arasındaki ilişkilerin incelenmesi ve her bir alt sürecin iş akışı üzerinde tanımlanması gereklidir.

Şekil 1'de iş akış sistemine entegre edilen yer sistem modelleri için tasarlanacak sistemde bulunması gereken bileşenler görülebilir. Sistem üç temel bileşenden oluşmaktadır. Bunlar hesaplama ortamı, tüm sistemin yönetilmesi için



Şekil 1. İş akış sisteminin genel yapısı ve bileşenleri

kullanılacak iş akışı uygulaması ve toplanan ön veri/köken bilgisinin kullanıcılara sunulması için geliştirilecek ağ veri geçididir. Ancak toplanan köken bilgisinin ağ üzerinden kullanıcılara sunulmasını sağlamak amacıyla kullanılan veri geçidinin tasarlanması ve geliştirilmesi bu çalışmanın kapsamı dışındadır.

Şekil 1'den görülebilen ve sistemin merkezinde bulunan iş akışı uygulaması araştırmacı ile diğer modelleme bileşenleri arasında bir çeşit soyutlaştırma katmanı işlevi görerek farklı teknolojilerin birbirleri ile uyumlu bir şekilde çalışmasını sağlamaktadır. İş akışı, tekrar eden birçok alt süreci yapısı içinde ve modüler bir şekilde tanımlanmasına izin vererek tüm bu süreçler için merkezi bir yönetim sistemi oluşturulmasını sağlar.

Tasarlanan iş akış ortamının bir diğer önemli görevi ise standartlaştırılmış bir şekilde ön veri ve köken bilgisinin toplamaktır. Toplanan bilgi çalışmanın daha sonraki aşamalarında hata ayık-

lama, sonuçların karşılaştırılması, model performansının geliştirilmesi ve belirli bir çalışmanın sonucunun doğrulanması veya tekrar edilmesi amacıyla kullanılabilir. Son yıllarda köken bilgisinin kullanılarak iş akışının tekrar üretilmesi üzerine yoğun çalışmalar mevcuttur.

Ön veri ve köken bilgisinin belirli bir standarda uymadan geliştirilmiş yer sistem modelinden toplanması ve bu modellerin iş akış sistemlerine entegre edilmesi oldukça karmaşık ve zor bir süreçtir. Bu konuda son zamanlarda yapılan çalışmalar yukarıda tanımlanan açığın kapatılması ve modellerin daha standart bir yapı altında toplanması üzerine yoğunlaşmıştır. Bu amaçla çerçeve yaklaşımı geliştirilmiş ve modellerin birbirleri ile olan arayüzlerinin standart bir yapıya kavuşturulması sağlanmaya çalışılmıştır. Bu sebeple iş akış ortamına entegre edilecek model veya model sisteminin çerçeve yaklaşımının avantajlarından yararlanılarak geliştirilmesi, sistemin farklı uygulamalar ile entegre edilmesini kolaylaştırmaktadır.

Yer sistem modellerinin iş akış ortamına aktarılmasında karşılaşılabilecek en önemli zorluklardan biri kullanılan modellerin çalışma sürelerinin çok uzun olmasıdır ve bu tip çalışmaların başlat, çalıştır ve incele modeli ile yapılması pek mümkün değildir. Bu problemin çözülmesi için tasarlanan iş akışı alt iş akışlarına ayrılarak birbirlerini izleyecek şekilde çalıştırılmaları sağlanabilir. Geliştirilecek iş akışı ortamı uzak hesaplama sistemine verilen bir iş bittiğinde bir sonraki iş akışını otomatik olarak çalıştıracak şekilde tasarlanması gereklidir. Şekil 2’de bu amaçla geliştirilen iş akışının temel yapısı görülebilir.

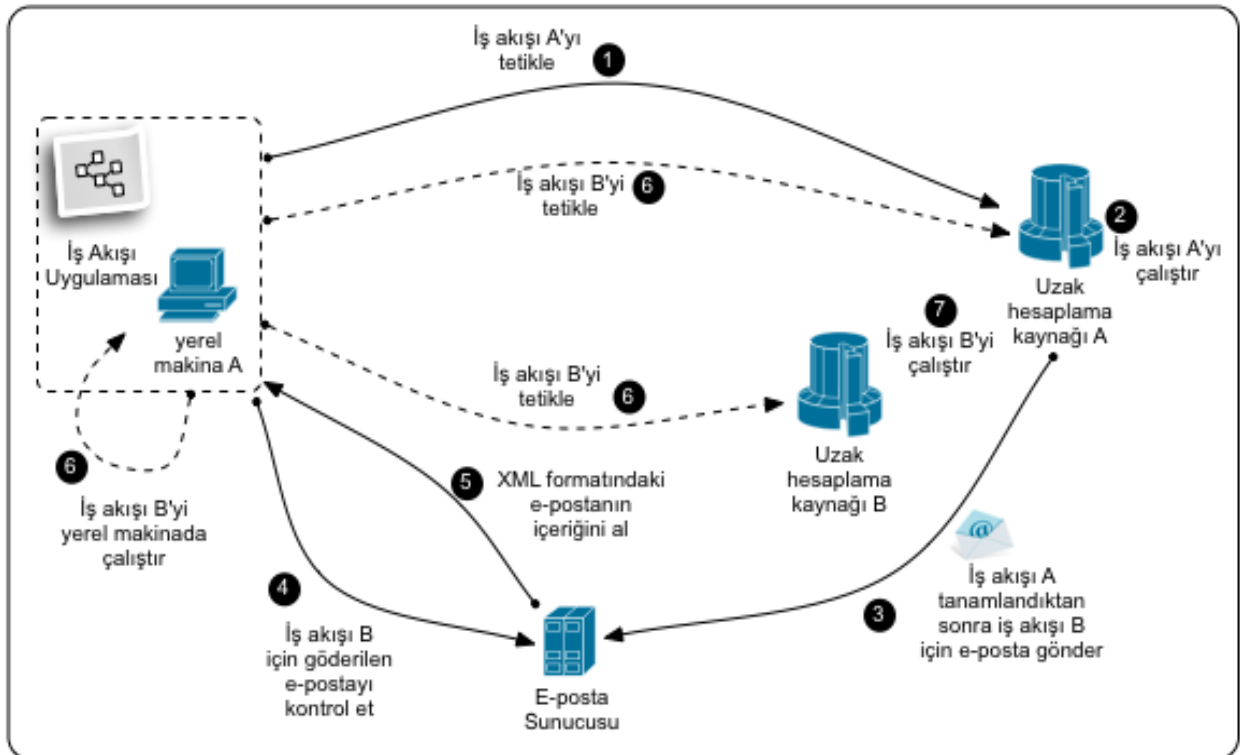
Önerilen metodolojide bu sistemin kurulması için biten işin daha önce tanımlanmış bir e-posta adresine mesaj atması (adım 3) ve bu mesajın içeriğini kontrol eden iş akışı uygulaması (adım 4 ve 5) tetiklenecek diğer iş akışı üzerindeki parametreleri değiştirerek çalıştırmaktadır.

İş akışı uygulaması: Kepler

Bir önceki bölümde tanımlanan metodolojinin en önemli parçası merkezi iş akışı bileşenidir. Literatürde birçok farklı bilimsel iş akışı uygulaması bulunmasına rağmen Kepler uygulaması

platform bağımsız, açık kaynak kodlu olması ve yer sistem bilimlerinde yapılan çalışmaların gereksinimlerini karşılayabilecek aktör/direktör yapısını temel alması nedeniyle seçilmiştir.

Kepler, Berkeley Üniversitesi tarafından geliştirilmiş Ptolemy II (Eker vd., 2003) iş akışı uygulamasının geliştirilmiş ve daha kolay kullanılabilir hale getirilmiş bir sürümüdür. Kepler’de kullanıcı bir süreci iş akış ortamına aktarmak için aktör adı verilen yapıları kullanır. Aktörler yapılarında bulunan girdi/çıkış portları ve parametreler ile belirli bir işi yapmak için özelleşmiş modüler bileşenlerdir ve Kepler uygulaması hazır birçok aktör ile birlikte gelmektedir. Hazır aktörler dışında kullanıcı Java programlama dili ile yeni aktörler tasarlayabilir. Üretilen iş akışının işleyişi direktör adı verilen özel aktörler yardımıyla kontrol edilir. Bu yapı kullanıcıya iş akışı içinde bulunan direktörü değiştirerek farklı bir hesaplama modeli (MoC) ile çalışma olanağı sağlar. Kepler içinde oluşturulan iş akışları ve aktör tanımları MoML (Lee ve Neuendorffer, 2000) adı verilen özel bir XML dosyası içinde saklanır.



Şekil 2. Yeni iş akışının otomatik olarak tetiklenmesi süreci

Kepler iş akışı uygulaması ile iç içe yuvalanmış bir şekilde hiyerarşik iş akışlarının tasarlanması mümkündür. Bu amaçla Kepler içindeki kompozit aktörler kullanılır. Bu özelleşmiş aktörler yapılarında kendi direktörü olan veya olmayan iş akışları tasarlanmasına izin vererek iş akışı tasarımının basitleştirilmesini ve kolay anlaşılmasını sağlar. Bu çalışma kapsamında farklı model bileşenlerini (atmosfer ve okyanus modelleri) tanımlayan süreçlerin iş akışına aktarılması amacıyla kompozit aktörlerden yararlanılmıştır.

Uygulama örneği: WRF ve ROMS

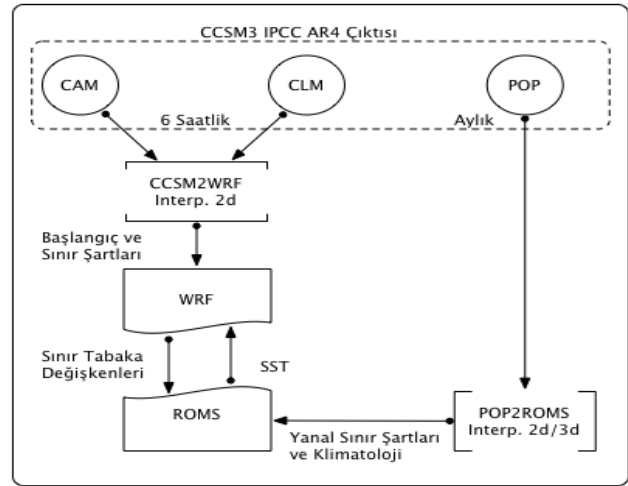
Bu bölüm, önceki bölümlerde tanımlanan metodolojinin sınanması için kuple edilmiş bir model sisteminin geliştirilmesi ve geliştirilen modelin iş akışının ortamına aktarılması hakkında bilgi vermeyi amaçlamaktadır.

Kuple model sisteminin geliştirilmesi için seçilen modeller bölgesel bir atmosfer modeli olan WRF (Janjic vd., 2001) ve yine bölgesel bir okyanus modeli olan ROMS (Shchepetkin vd., 2004) modelleridir. Bu test uygulamasında amaç geliştirilen kuple model sistemi üzerinde hassasiyet analizleri yapmak değil, geliştirilen model sistemini, tanımlanan metodoloji çerçevesinde iş akışı ortamına aktarmaktır. Bu nedenle modelleme sistemi ile uzun süreli iklimsel simulasyonlar yapmak yerine sadece sistemin düzgün bir şekilde çalıştığını göstermesi için iklimsel çalışmalar bakımından kısa süreli sayılabilecek bir benzetim tasarlanmıştır.

Kuple model sisteminin içerdiği adımlar Şekil 3'te görülebilir. Öncelikli olarak CCSM küresel modeli çıktıları (IPCC SRES 20C) kullanılarak ROMS ve WRF için geliştirilen programlar yardımıyla gerekli olan başlangıç ve sınır koşullarına ait dosyalar üretilmiştir. Model sisteminin bileşenlerinden birini oluşturan atmosfer modelinden okyanus modeline sınır tabaka değişkenleri, diğer bileşen olan okyanus modelinden ise atmosfer modeline deniz yüzeyi sıcaklığı (SST) transfer edilmektedir.

Her iki model Akdeniz bölgesini kapsayacak şekilde ayarlanmıştır. Ancak bu uygulamada atmosfer modeli okyanus modelinden daha büyük bir alanı kapsamaktadır ve her iki modelin

grid noktaları aynı olmadığından interpolasyon yolu ile gridler arasında dönüşüm yapılmalıdır. Bu amaçla gridler arasında interpolasyon ağırlıkları SCRIP (Jones 1999) uygulaması yardımıyla hesaplanmış ve noktalar arasındaki uzaklığın temel alındığı bir interpolasyon yöntemi kullanılmıştır. ESMF uygulaması bu ağırlık matrisi yardımıyla farklı iki yer sistem modeli gridi arasında seyrek matris çarpımı yöntemi ile alan değişkenlerini gridler arasında transfer etmektedir.

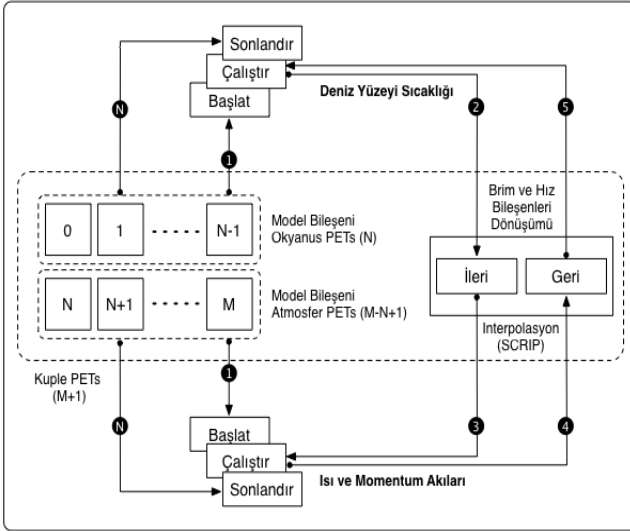


Şekil 3. Test uygulamasının genel yapısı

Şekil 4'te ESMF kütüphanesi kullanılarak kuple edilmiş modellerin genel yapısı görülebilir. Şekilden görülebileceği gibi kuple bileşeni model gridleri arasında interpolasyon yapmak için ileri (okyanus-atmosfer) ve geri (atmosfer-okyanus) olarak isimlendirilen iki farklı yönde çalışmaktadır ve her bir çalışma yönü kendisine ait değişik bir ağırlık matrisi kullanmaktadır. Alan değişkenlerinin model bileşenleri arasında aktarılmasından ESMF sorumludur. ESMF bu amaçla durum nesnelere adı verilen yapıları kullanır. İçinde alan değişkenleri barındıran bu yapılar ESMF tarafından veri alışverişi yapmak amacıyla tasarlanmıştır.

Geliştirilen model sisteminden köken bilgisinin toplanması amacıyla her bir modele ait kurulum yapı ve betikleri ORNL/NCSU tarafından geliştirilen Python betiklerini kullanacak şekilde değiştirilmiştir. Bu şekilde sistem köken bilgisi kurulum anında kuple model sisteminden otomatik olarak toplanmaktadır. Veri köken bilgisinin toplanması amacıyla ise geliştirilen model

sisteminde ESMF özellik yapılarının kullanılabilmesi şeklinde değişiklik yapılmıştır. Bu değişiklik ile model sistemi içinde alan değişkeni ve model bileşeni bazında ön veri eklenmiştir.



Şekil 4. Kupule model sisteminin yapısı

Geliştirilen iş akışı

Bu bölüm kupule edilmiş model sistemini iş akışı ortamına aktarılması için geliştirilen aktörler ve iş akışı hakkında bilgi vermeyi amaçlamaktadır. Kupule model sisteminin iş akışı sistemine aktarılmasını kolaylaştırmak amacıyla öncelikle her bir alt modele ait iş akışının tekil olarak tasarlanmasına karar verilmiştir. Tekil olarak tasarlanan iş akışları daha sonra birleştirilerek kupule model sistemini tanımlayan iş akışı oluşturulmuştur. Şekil 5'te iş akışına aktarılması planlanan tüm adımlar görülebilir.

Mantıksal iş akışı incelendiğinde birbirleri ile ilişkili iki grup iş akışı olduğu görülmektedir. Birinci grup atmosfer modeli ile ilgili olan süreçleri, ikinci grup ise okyanus modeli ile ilgili olan süreçleri içermektedir. Şekilden de görülebileceği gibi okyanus modeli ile ilgili süreçler atmosfer modeli ile ilgili süreçlere bağımlıdır.

İş akışının Kepler uygulaması kullanılarak tasarlanması ve çalıştırılması için hesaplama kaynağı olarak bir küme sistem olan NCAR'a ait Bluefire sunucusu kullanılmıştır. Bu nedenle uzak sunucu ile olan iletişim tek kullanımlık şifre ve SSH protokolü ile sağlanmaktadır. Bu amaçla uzak makina üzerinde komut çalıştırma ve veri

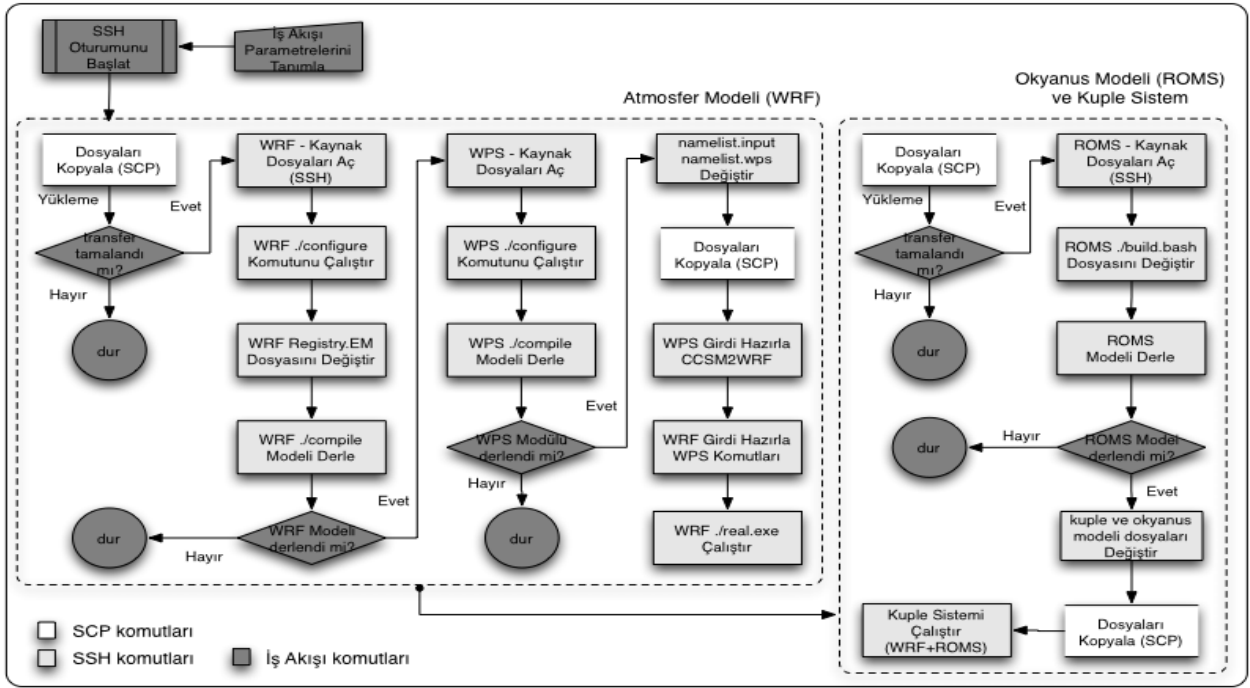
transferi için Kepler içinde bulunan SSH aktörleri kullanılmıştır. Ancak hesaplama aktörlerinin değiştirilmesi ile hazırlanan iş akışı farklı bir hesaplama ortamında (grid yapısı) da çalışabilecek şekilde değiştirilebilmektedir.

Kepler uygulaması kullanılarak her bir model bileşenine ait ayar dosyalarının değiştirilmesi amacıyla *WRFEnvModify* ve *ROMSEnvModify* olarak adlandırılan iki aktör geliştirilmiştir. Bu özel aktörler modellere ait parametre dosyalarını istenilen şekilde değiştirerek farklı model benzetimlerinin yaratılmasını olanaklı hale getirmektedir. Şekil 5'te görülen mantıksal iş akışının Kepler sistemine aktarılmış hali Şekil 6'da görülebilir.

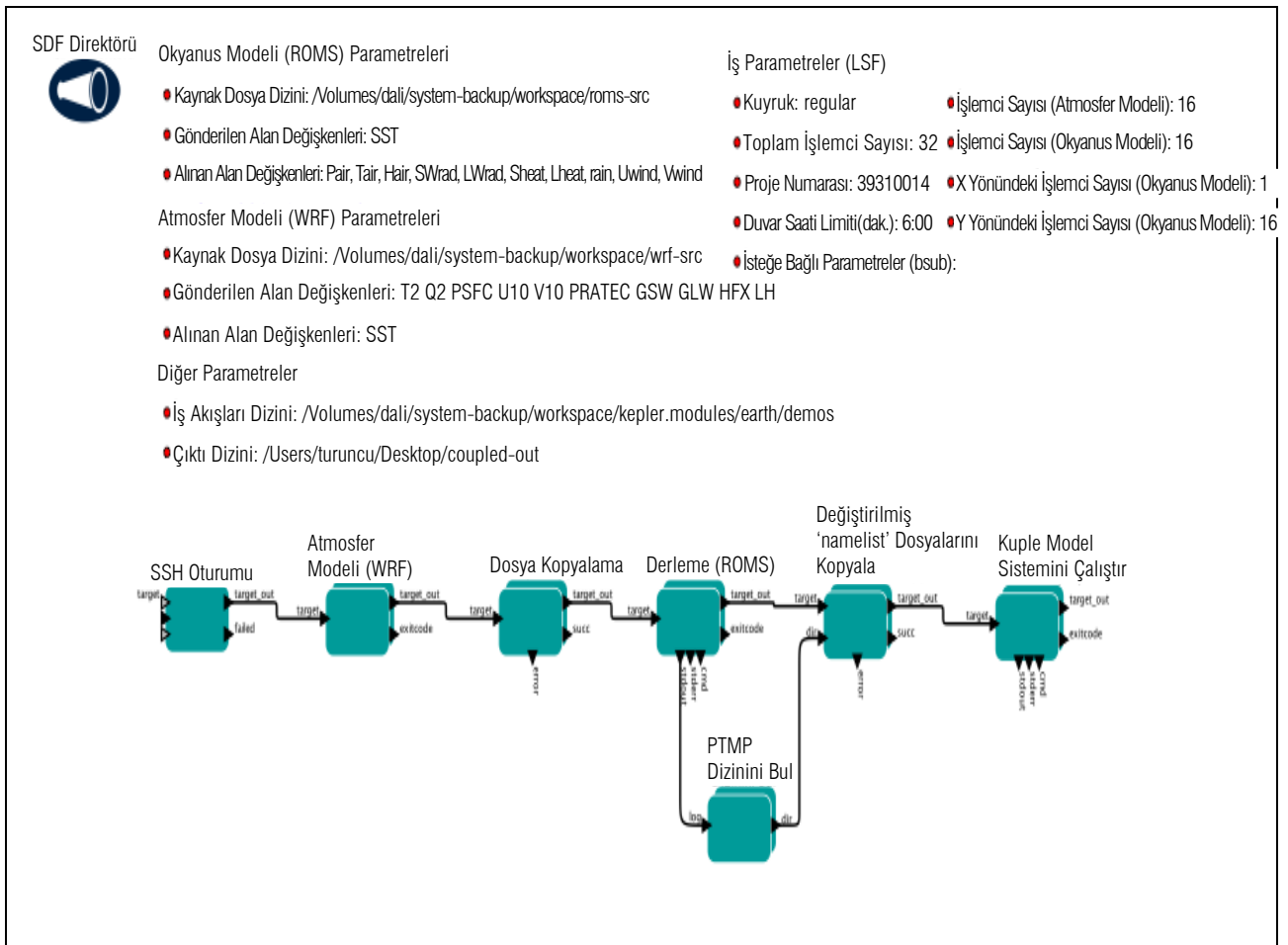
Şekildeki iş akışı incelendiğinde iş akışının işleyişini değiştiren temel parametrelerin en üst seviye iş akışında tanımlandığı görülebilir. Bu parametreler arasında model kaynak kodlarının bulunduğu dizinler, uzak sistemde bulunan ve sisteme verilen işleri düzenleyen yazılma (LSF) verilecek parametreler ve model bileşenleri arasında değiştirilecek olan alan değişkenleri sayılabılır.

Kullanılan kompozit aktörler ile her bir modele ait alt süreçler farklı iş akışı seviyesinde tanımlanarak daha kolay anlaşılır bir yapı kurulabilir. Bu amaçla atmosfer modeline ait iş akışı bir kompozit aktör içerisinde kupule model sistemini oluşturan iş akışına entegre edilmiştir.

Tasarlanan iş akışı ile otomatik olarak toplanan köken bilgisi temel olarak sistem ve veri köken bilgilerinden oluşmaktadır. Sistem köken bilgisi uzak sistemde çalıştırılan komutlar için gerekli çevre değişkenlerini tanımlayan örtü betik ile ve derleme aşamasında kullanılan Python betikleri ile toplanmaktadır. Bu betikler ile toplanan köken bilgisine örnek olması açısından Şekil 7'deki ROMS modeline ait dosya incelenebilir. XML çıktısı temel olarak üç grup bilgi içermektedir. Bunlar çalıştırılabilir dosya ile ilgili olan bilgiler, modelin derlenmesi sırasında kullanılan derleyici ile ilişkili bilgiler ve tanımlanmış olan çevre değişkenlerinin listesi olarak sıralanabilir. Tüm bu bilgiler model sisteminin çalıştırıldığı uzak sistem üzerindeki şartların yaratılması için kullanılabilir.



Şekil 5. Kuple model sistemi mantıksal iş akışı tasarımı



Şekil 6. Kuple model sistemini (WRF&ROMS) tanımlayan iş akışı

```
<?xml version="1.0" encoding="UTF-8"?>
<system_provenance>
  <run>
    <component>rev_382</component>
    <executable>/ptmp/turunco/case/oceanM</executable>
    <checksum>d2c898797c9f6aaa2ef678b0bdd2a2a9</checksum>
    <user>turunco</user>
    <time>09-09-28 15:35:41</time>
  </run>
  <compilers>
    <compiler>
      <property name="command" value="mpxf95_r"/>
      <property name="description" value="IBM XL Fortran Enterprise Edition for AIX, V11.1"/>
      <property name="flags" value="-qsuffix=f=f90 -qmaxmem=-1 -qarch=pwr4 -qtune=pwr4 -q64 -O3 -qstrict -I/ptmp/turunco/esmf_3.1.0rp2/mod/modO/AIX.default.64.mpi.default -I/ptmp/turunco/esmf_3.1.0rp2/src/include"/>
      <property name="language" value="Fortan"/>
      <property name="version" value="11.01.0000.0007"/>
    </compiler>
  </compilers>
  <environment>
    ...
    ...
    <variable name="ESMF_ABI" value="64"/>
    <variable name="ESMF_BOPT" value="O"/>
    <variable name="ESMF_COMM" value="mpi"/>
    <variable name="ESMF_COMPILER" value="default"/>
    <variable name="ESMF_DIR" value="/ptmp/turunco/esmf_3.1.0rp2"/>
    <variable name="ESMF_FORTRANSYMBOLS" value="default"/>
    ...
    ...
  </environment>
```

Şekil 7. ROMS modeli örnek sistem köken bilgisi çıktısı

ESMF özellik nesnelere ve buna bağlı metodlar ile toplanan veri köken bilgisi model bileşeni ve değişkenler ile ilgili modelden gelen ön veri bilgilerini içermektedir (Şekil 8).

Dosya iki grup veri içermektedir. Birinci grup, model bileşeninin hangi disipline ait olduğu, hangi merkez ve hangi programlama dili kullanılarak yazıldığı vb. bilgiler içermektedir. İkinci grup bilgiler arasında ise değişkenlere ait isim, standart isim, tanımlama ve birim gibi bilgiler içermektedir. Değişken tanımlamaları içinde bulunan “import” ve “export” tanımları model değişkeninin ilgili bileşen tarafından üretilip üretilmediğini belirtmektedir ve bu tanımlar ESMF durum nesnelere tarafından durum nesnesinin tipine göre otomatik olarak üretilmektedir. Örneğin Şekil 8’deki XML dosyasında bulunan SST değişkeni okyanus modeli ROMS’tan atmosfer modeli WRF’ye transfer edilen bir değişkendir. Bu şekilde değişkenler ile ilgili detaylı bilgi tutulmuş olur.

```
<model_component name="WRF" full_name="Weather Research and Forecasting" version="Standard 3.1.1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.esmf.ucar.edu file:/esmf_model_component.xsd"
  xmlns="http://www.esmf.ucar.edu">
  <discipline_set>
    <discipline name="Atmosphere" />
  </discipline_set>
  <physical_domain_set>
    <physical_domain name="Earth system" />
  </physical_domain_set>
  <agency_set>
    <agency name="NCAR (National Center for Atmospheric Research)" />
  </agency_set>
  <institution_set>
    <institution name="MMM (Mesoscale and Microscale Meteorology Division)" />
  </institution_set>
  <author_set>
    <author name="M3 Team" />
  </author_set>
  <coding_language_set>
    <coding_language name="Mixed Fortran 90 and C" />
  </coding_language_set>
  <model_component_framework_set>
    <model_component_framework name="ESMF (Earth System Modeling Framework)" />
  </model_component_framework_set>
  <variable_set>
    <variable name="SST"
      standard_name="sea_surface_temperature"
      long_name="Sea surface temperature"
      units="Kelvin"
      import="true"
      export="false" />
    ...
    ...
  </variable_set>
```

Şekil 8. WRF modeli örnek veri köken bilgisi çıktısı

Sonuçlar

Önceki bölümde detaylı olarak anlatılan sına uygulamalarının iş akışına entegre edilerek köken bilgisinin otomatik olarak toplanması kendi kendisini tanımlayan modellerin geliştirilmesi için önemli bir örnektir. Yer sistem bilimlerinde sıklıkla kullanılan ROMS ve WRF gibi iki modelin iş akış sistemine entegre edilmesi ile bu tip çalışmalar için temel teşkil edebilecek yeni bir metodoloji geliştirilmiş ve yapı kurulmuştur.

Önceki bölümlerde değinilen problemlerin çözümüne yönelik olarak geliştirilmiş olan iş akışı ve çerçeve yaklaşımlarının birlikte kullanılabilirliği bu çalışma ile sinanmış ve geliştirilen iş akışı uygulaması ile yer sistem bilimlerinde sıkça kullanılan süreçlerin basitleştirilerek ortalama bir kullanıcı seviyesine indirgenmesi sağlanmıştır.

Kaynaklar

- Bettencourt, M. T., (2002). Distributed Model Coupling Framework. 11th IEEE International Symposium on High-Performance Distributed Computing (HPDC-11 '02), 284, Edinburgh, Scotland.
- Bowers, S. and Ascher, B., (2005). Actor-oriented design of scientific workflows, *In 24st Intl. Conference on Conceptual Modeling*, Springer, 369–384.
- Bowers, S., McPhillips, T.M., Ludäscher, B., Cohen, S. and Davidson, S.B., (2006). A Model for User-Oriented Data Provenance in Pipelined Scientific Workflows. International Provenance and Annotation Workshop, IPAW 2006. Chicago, IL, USA.
- Collins, N., Theurich, G., Deluca, C., Suarez, M., Trayanov, A., Balaji, V., Li, P., Yang, W., Hill, C. and Da Silva, A., (2005). Design and Implementation of Components in the Earth System Modeling Framework, *International Journal of High Performance Computing Applications*, **19**, 3, 341–350.
- Craig, A.P., Jacob, R., Kauffman, B., Bettge, T., Larson, J., Ong, E., Ding, C. and He, Y., (2005). CPL6: The New Extensible, High Performance Parallel Coupler for the Community Climate System Model, *International Journal of High Performance Computing Applications*, **19**, 3, 309–327.
- Eker, J. W. Janneck, E. A. Lee, J. Liu, X. Liu, J. Ludvig, S. Neuendorffer, S. Sachs, Y. Xiong, (2003). Taming Heterogeneity - the Ptolemy Approach, *In Proceedings of the IEEE*, **91**, 1, 127-144.
- Hill, C., DeLuca, C., Balaji, V., Suarez, M. and da Silva, A., (2004). The Architecture of the Earth System Modeling Framework, *Computing in Science and Engineering*, **6**, 1, 18–28.
- Janjic, Z. I., J. P. Gerrity, Jr. and S. Nickovic, (2001). An Alternative Approach to Nonhydrostatic Modeling, *Monthly Weather Review*, **129**, 1164-1178.
- Jones, P. W., (1999). First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates, *Monthly Weather Review*, **127**, 2204–2210.
- Larson, J., Jacob, R. and Ong, E., (2005). The Model Coupling Toolkit: A New Fortran90 Toolkit for Building Multiphysics Parallel Coupled Models, *International Journal of High Performance Computing Applications*, **19**, 3, 277–292.
- Lee, E.A. and Neuendorffer, S., (2000). MoML - A Modeling Markup Language in XML - Version 0.4, Technical report, University of California at Berkeley, USA.
- Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M., Lee, E.A., Tao, J. and Zhao, Y., (2006). Scientific workflow management and the Kepler system, *Concurr. Comput. : Pract. Exper.*, **18**, 10, 1039-1065.
- Majithia, S., Shields, M., Taylor, I. and Wang, I., (2004). Triana: a graphical Web service composition and execution toolkit, *Web Services, Proceedings*, IEEE International Conference on, 514–521.
- Oinn, T., Addis, M., Ferris, J., Marvin, D., Carver, T., Pocock, M.R. Wipat, A., (2004). Taverna: A tool for the composition and enactment of bioinformatics workflows, *Bioinformatics*, **20**, 17, 3045-3054.
- Plale, B., Gannon, D., Brotzge, J., Droegemeier, K., Kurose, J., McLaughlin, D., Wilhelmson, R., Graves, S., Ramamurthy, M., Clark, R.D., Yalda, S., Reed, D.A., Joseph, E. and Chandrasekar, V., (2006). CASA and LEAD: Adaptive Cyberinfrastructure for Real-Time Multiscale Weather Forecasting, *Computer*, **39**, 11, 56–64.
- Shchepetkin, A.F., J.C. McWilliams, (2004). The Regional Oceanic Modeling System: A split-explicit, free-surface, topography-following-coordinate ocean model, *Ocean Modelling*, **9**, 347-404.