

SISTEMA MULTI-ESCALA DE PREVISÃO EM TEMPO REAL DA DINÂMICA ESTUARINA E COSTEIRA: DESAFIOS PARA A OPERACIONALIZAÇÃO EM AMBIENTE *CLOUD* E HPC

Anabela OLIVEIRA¹; João ROGEIRO²; Alberto AZEVEDO³; André B. FORTUNATO⁴;
Ricardo TAVARES DA COSTA⁵; Marta RODRIGUES⁶; Kai LI⁷; João Paulo MARTINS⁸;
Mário DAVID⁹; João PINA¹⁰; Jorge GOMES¹¹

RESUMO

Os sistemas de previsão em tempo real (SPTR) fazem previsões a escalas de tempo curtas, por integração de modelos e dados de campo, constituindo ferramentas de suporte à gestão costeira. Os SPTR devem utilizar modelos numéricos de grande precisão, que requerem ambientes de computação exigentes. O LNEC opera vários SPTR, que utilizam a Infraestrutura Nacional de Computação Distribuída. A computação *cloud*, que permite a utilização de máquinas virtuais utilizando servidores compartilhados e interligados através da Internet, surge agora como um paradigma alternativo. Os seus recursos podem ser disponibilizados como um serviço de computação, sendo uma alternativa para o processamento exigente dos SPTR. No entanto, não existe uma avaliação detalhada da *cloud* por comparação com estratégias convencionais para processamento dos SPTR. Apresenta-se aqui uma análise comparativa da performance de vários SPTR utilizando três abordagens distintas: *workstations*, *Clusters* e *Cloud*. Esta comparação foi elaborada em três aplicações de SPTR com diferentes exigências de computação: circulação na Ria de Aveiro; circulação acoplada a indicadores de

¹ Eng.^a Civil, Doutora em Eng.^{ia} do Ambiente, LNEC, Av. do Brasil 101, 1700-066 Lisboa, aoliveira@lnec.pt

² Eng.^o Informático, Bolseiro de Investigação, LNEC, Av. do Brasil 101, 1700-066 Lisboa, jrogeiro@lnec.pt

³ Oceanógrafo físico, Doutor em Ciências Geofísicas e da Geoinformação, LNEC, Av. do Brasil 101, 1700-066 Lisboa, aazevedo@lnec.pt

⁴ Eng.^o Civil, Doutor em Eng.^{ia} do Ambiente, LNEC, Av. do Brasil 101, 1700-066 Lisboa, afortunato@lnec.pt

⁵ Eng.^o do Ambiente, Doutorando em Engenharia Civil, LNEC, Av. do Brasil 101, 1700-066 Lisboa, rcosta@lnec.pt

⁶ Eng.^a do Ambiente, Doutora em Biologia, LNEC, Av. do Brasil 101, 1700-066 Lisboa, mfrdrigues@lnec.pt

⁷ Oceanógrafo, Bolseiro de Investigação, LNEC, Av. do Brasil 101, 1700-066 Lisboa, likai@foxmail.com

⁸ Lic. Física Tecnológica, Investigador do LIP

, LIP, Av. Elias Garcia 14 - 1º, 1000-149 Lisboa, martinsj@lip.pt

⁹ Doutor em Física, LIP, Av. Elias Garcia 14 - 1º, 1000-149 Lisboa, david@lip.pt

¹⁰ Lic. Física Tecnológica, Doutor em Física, LIP, Av. Elias Garcia 14 - 1º, 1000-149 Lisboa, jpina@lip.pt

¹¹ Doutor em Física, LIP, Av. Elias Garcia 14 - 1º, 1000-149 Lisboa, Jorge@lip.pt

contaminação fecal no estuário do Tejo; e circulação acoplada de ondas e correntes no mesmo estuário.

A análise de performance foi efetuada nos três ambientes utilizando números crescentes de processadores. Para a gama de processadores comuns, as duas alternativas conduzem a tempos computacionais semelhantes, mostrando que a *cloud* é uma opção viável e atractiva.

No entanto, as malhas de cálculo para a computação multi-escala (desde o oceano até à área portuária) requerem um número óptimo de processadores tipicamente superior. A evolução eficaz dos SPTR, baseados atualmente em estratégias de computação de elevada performance, para arquiteturas de *cloud* poderá assim requerer a integração de um maior número desses recursos através de federação. A performance destas simulações deverá, no entanto, ser analisada para assegurar que o tempo de comunicação entre recursos *cloud* não afeta significativamente o tempo necessário para as previsões.

Palavras-chave: previsão em tempo real; computação paralela; *cluster*; *cloud*; estuários.

1. INTRODUÇÃO

Os estuários são sistemas de elevada produtividade biológica, suportando múltiplas atividades antropogénicas e providenciando vários serviços de ecossistemas. O aumento previsto da pressão humana nas zonas costeiras assim como os impactos devidos às alterações climáticas podem aumentar o risco na manutenção destes serviços e até alterar a dinâmica dos seus ecossistemas. A capacidade de simular e prever a dinâmica estuarina e costeira é assim da maior importância para suportar a gestão otimizada dos seus recursos.

Os sistemas de previsão em tempo real (SPTR) contribuem para este objetivo, produzindo previsões das grandezas relevantes a escalas temporais curtas, através da integração de modelos numéricos e redes de monitorização em tempo real. A confiabilidade destas previsões é, no entanto, limitada pela precisão dos modelos numéricos e pelas condições da sua aplicação. Os modelos de malhas não estruturadas têm sido usados há várias décadas para simular as zonas costeiras (Dietrich *et al.*, 2012), mas a utilização das escalas espaciais e temporais adequadas à modelação multi-escala e a estudos de alterações climáticas é frequentemente limitada pelos recursos computacionais disponíveis. O custo computacional é, em particular, uma das maiores limitações no uso de modelos de elevada precisão em SPTR em zonas costeiras (Rodrigues *et al.*, 2013).

Há mais de uma década que os SPTR do LNEC são processados em *workstations*, procurando a robustez e o controle de qualidade através da comparação automática com dados em tempo real e redundância de simulações em recursos computacionais distintos. Numa fase posterior, o acesso a recursos de computação e de armazenamento distribuídos baseados em paradigmas de *cluster/grid* permitiram melhorar o tempo de resposta ao utilizador, existindo actualmente vários SPTR baseados nestas infraestruturas (Davies *et al.*, 2006, Oliveira *et al.*, 2014).

No entanto, a necessidade de recursos continua aumentar à medida que o conhecimento sobre os processos e a análise multi-escala vai crescendo (e.g. uso de várias simulações para cada compartimento e a sua integração em *ensembles*, para forçamento dos compartimentos a jusante). Adicionalmente, é necessário procurar soluções que permitam operacionalizar os SPTR nas entidades gestoras, sem necessidade de equipamento e de pessoal especializado para a manutenção da infraestrutura computacional.

A *cloud surge* assim como uma alternativa flexível para abordar os dois desafios, graças à sua capacidade de angariar recursos à medida das necessidades e aos serviços para facilitar a sua utilização (Buyya *et al.*, 2008). No entanto, o uso da *cloud* para computação intensiva e complexa é ainda recente e não foi ainda estudado em detalhe. O tempo de comunicação entre nós em computação paralela, a segurança dos dados e a qualidade do serviço (em particular a garantia de entrega atempada das previsões) necessita de análise antes de migrar os serviços operacionais existentes ou criar novos serviços para este novo tipo de infraestrutura.

Dado que cada recurso computacional tem vantagens e desvantagens dependendo da aplicação específica, a análise aqui apresentada é sustentada por três usos distintos de um SPTR. Existem apenas algumas comparações entre plataformas alternativas (e.g. Glenis *et al.*, 2013) e nenhuma é dedicada a SPTR para circulação e qualidade da água em zonas costeiras.

O SPTR utilizado neste trabalho é parte da infraestrutura WIFF – *Water Information Forecast Framework* (Oliveira *et al.*, 2014, Figura 1), e é baseado na aplicação de uma plataforma genérica, adaptável a qualquer localização geográfica, que foi customizada no LNEC para aplicações costeiras e suas interfaces (Jesus *et al.*, 2012). O sistema integra um conjunto de modelos que é utilizado diariamente em ambientes de elevada performance (HPC) para simular níveis, correntes, agitação marítima, salinidade, temperatura da água e várias variáveis de qualidade da água (indicadores de contaminação fecal e derrames de hidrocarbonetos).

Esta comunicação apresenta uma análise de performance de vários ambientes computacionais para o funcionamento operacional de SPTR, com vista a determinar a exequibilidade da *cloud* para este fim. Esta análise é efetuada em três ambientes distintos: *workstations*, nós *de clusters HPC* e *cloud*, aplicada a três SPTR distintos, baseados em processamento paralelo e distribuído recorrendo a MPI: circulação na Ria de Aveiro; circulação acoplada a indicadores de contaminação fecal no estuário do Tejo; e circulação acoplada de ondas e correntes no mesmo estuário. A análise é focada no tempo necessário para disponibilizar os resultados da previsão em tempo real e no número óptimo de processadores para cada aplicação.

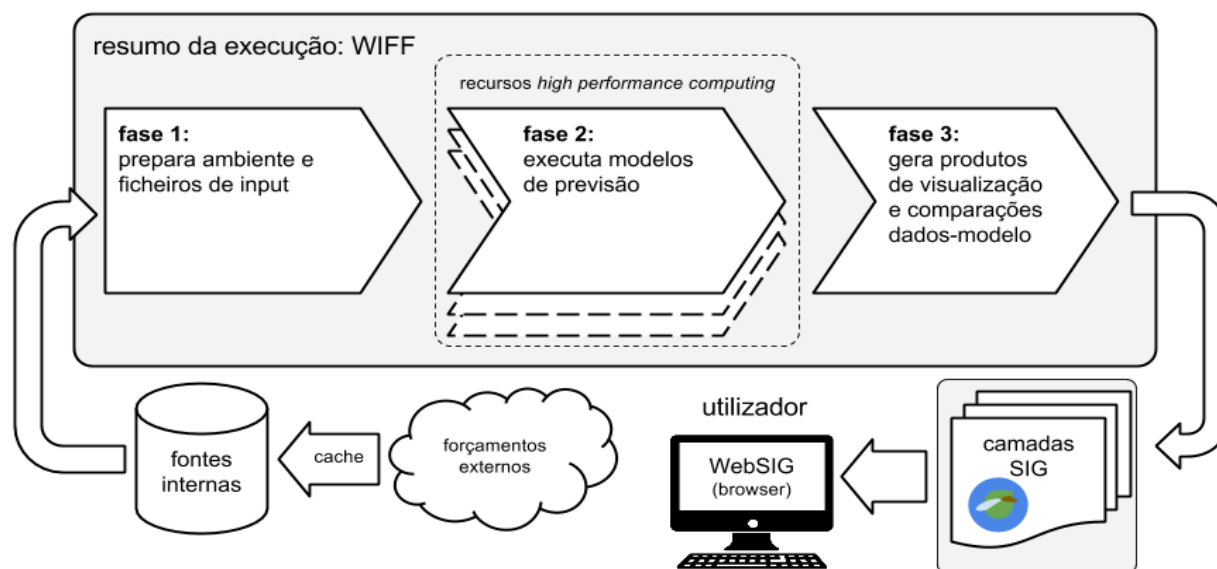


Figura 1. Workflow de um sistema de previsão em tempo real integrado na plataforma WIFF.

2. MODELAÇÃO MULTI-ESCALA DE ELEVADA PRECISÃO DA DINÂMICA DA COSTA PORTUGUESA

A simulação da dinâmica estuarina e costeira envolve necessariamente diferentes escalas espaciais. Por um lado, os diferentes processos físicos são diretamente afetados pelas geometrias dos sistemas, como margens, canais e obras costeiras. A correta representação destas geometrias exige resoluções espaciais da ordem da dezena de metros ou inferior. Simultaneamente, o escoamento é forçado por fenómenos de larga escala (centenas ou milhares de quilómetros), como a geração e propagação de marés, sobre-elevações de origem meteorológica e agitação marítima. A consideração simultânea das diferentes escalas pode ser conseguida através de sistemas de malhas encaixadas, de malhas não-estruturadas, que cobrem diferentes escalas em simultâneo, ou uma combinação das duas abordagens. Em qualquer caso, o esforço computacional é significativo, obrigando à utilização de meios de cálculo potentes, e de modelos paralelizados para tirar partido destes meios.

Os fenómenos de interesse são também progressivamente mais complexos, exigindo modelos mais sofisticados e computacionalmente exigentes. A crescente exigência sobre a precisão dos modelos requer o acoplamento de modelos para diferentes processos, como a circulação e a agitação marítima. Só através deste acoplamento é possível ter em conta as interações entre os processos (e.g., geração de correntes pela rebentação, refração das ondas pelas correntes). Da mesma forma, os modelos de qualidade da água incluem um número crescente de variáveis químicas e biológicas. Também esta tendência cria exigências elevadas nos meios

computacionais, sobretudo em modelos tridimensionais.

Os modelos utilizados nos SPTR do LNEC ilustram estas tendências (Figura 2). Simulam-se simultaneamente escalas de milhares de quilómetros (e.g., geração e propagação de marés e sobre-elevações no Atlântico Norte) e chegam-se a resoluções da ordem de 1 m (e.g., para simular descargas de águas residuais no estuário do Tejo), utilizam-se modelos acoplados (e.g. ondas e correntes), e representam-se processos de forma detalhada (e.g., trocas de óleo entre a coluna de água e o fundo em modelos de derrames de hidrocarbonetos).

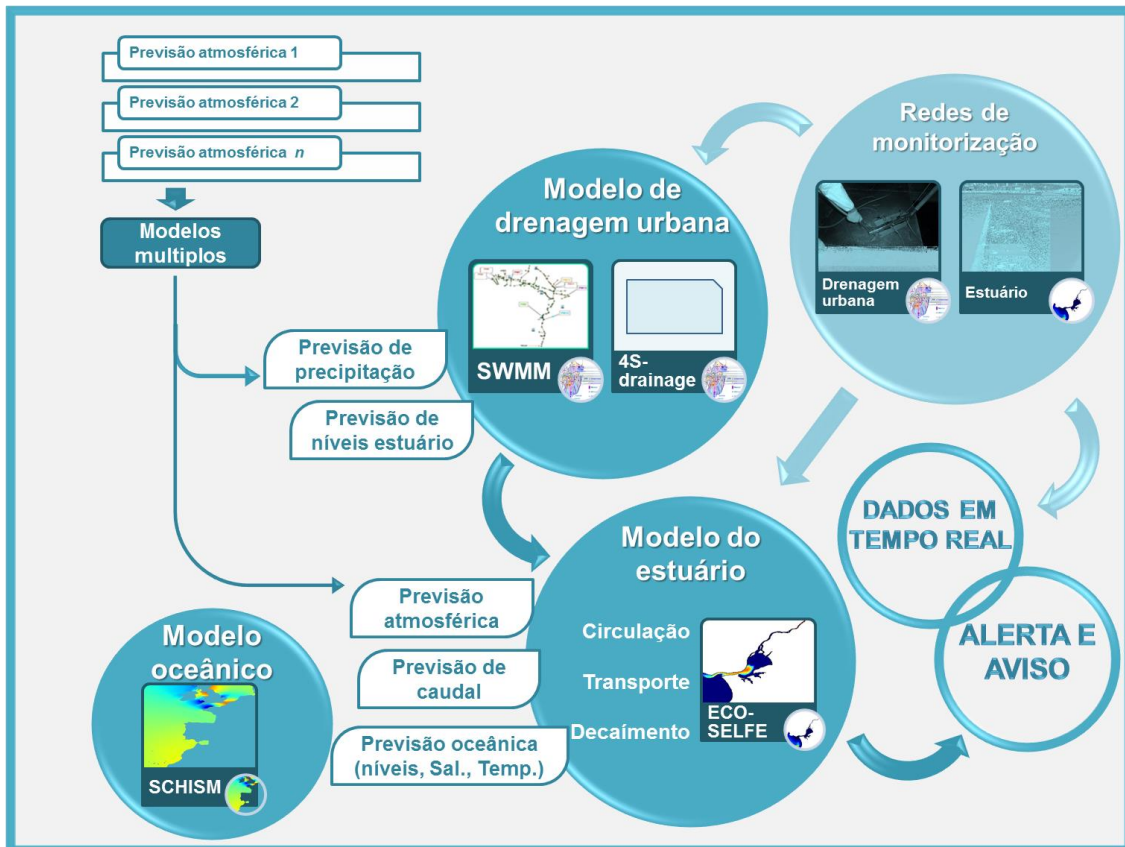


Figura 2. Modelação multi-escala da costa Portuguesa: da escala regional ao detalhe urbano. Exemplo do SPTR do estuário do Tejo.

3. SISTEMAS DE PREVISÃO EM TEMPO REAL EM ZONAS COSTEIRAS

A capacidade de simular e prever a dinâmica dos meios aquáticos é essencial para avaliar os impactos sociais, económicos e ecológicos das intervenções humanas e da variabilidade e alterações climáticas nestes ambientes. Os SPTR são um elemento chave neste processo (Werner *et al.*, 2013, Oliveira *et al.*, 2014), através da integração de redes de monitorização, modelos numéricos, recursos computacionais avançados para computação paralela ou

distribuída, para produção de previsões a escalas de tempo curtas (dias) e análise automática de resultados, e sistemas de informação Web. Estes sistemas constituem ferramentas habituais de suporte à gestão costeira e portuária para análise de grandezas como a agitação marítima, os níveis, as velocidades, a salinidade e temperatura, assim como quantidades relevantes para o controle da qualidade da água.

Os SPTR têm evoluído ao longo dos anos, quer ao nível dos processos simulados, quer ao nível das funcionalidades disponíveis. Os SPTR de previsão da circulação são já ferramentas comuns (e.g. http://www.opc.ncep.noaa.gov/Current_fcsts.shtml, <http://ariel.lnec.pt>). Para a previsão da qualidade da água, estes sistemas são ainda recentes (e.g. Rodrigues *et al.*, 2013, Oliveira *et al.*, 2014) e muitos desafios persistem devido à complexidade da integração entre os modelos em cascata, com diferentes escalas temporais e espaciais.

O custo computacional é uma das maiores limitações no uso de modelos de elevada precisão em SPTR nas zonas costeiras (Rodrigues *et al.*, 2013). Os avanços recentes em computação de elevada performance (Costa *et al.*, 2009) criaram as condições para o uso eficaz destes modelos em SPTR. O uso de modelos em modo paralelo abriu ainda caminho à evolução dos SPTR da hidrodinâmica para a circulação acoplada ondas-correntes (Fortunato *et al.*, 2015) e a qualidade da água (Rodrigues *et al.*, 2013), e à sua operacionalização diária através do uso de recursos computacionais avançados em *clusters*, sendo as previsões disponibilizadas on-line como serviços para as entidades gestoras (Oliveira *et al.*, 2014).

O balanço entre a necessidade de providenciar alertas em tempo útil (para evitar danos materiais e potenciais problemas de saúde pública) e a melhor resolução espacial e temporal para previsões confiáveis da circulação e qualidade da água motivou a presente procura de infraestruturas alternativas para operar os SPTR. Para além da necessidade de representar os processos às escalas adequadas, os requisitos computacionais para os SPTR incluem ainda os associados à modelação multi-escala em cadeia, desde o oceano à escala portuária ou urbana, e o tempo de comunicação entre recursos computacionais, de localização distribuída.

Neste trabalho são utilizados três SPTR em operação no LNEC para suportar a análise comparativa de performance e a identificação das vantagens e desvantagens das várias alternativas disponíveis atualmente para computação *HPC*.

4. PARADIGMAS ATUAIS PARA COMPUTAÇÃO HPC E RECURSOS NACIONAIS EXISTENTES

Os avanços recentes em infraestruturas computacionais têm promovido um avanço considerável em muitas áreas que são suportadas pela modelação numérica. Em consequência, a disponibilidade destes recursos tem permitido estudos de modelação com complexidade e detalhe crescentes, que criam novos desafios e necessidades de recursos computacionais. Simultaneamente, estes avanços científicos devem ser disponibilizados aos utilizadores finais, de forma a assegurar atividades de gestão do património e do meio ambiente devidamente sustentadas no melhor conhecimento disponível. No entanto, esta disponibilização cria algumas dificuldades operacionais às entidades gestoras, já que a manutenção destas infraestruturas computacionais requer pessoal especializado.

Tradicionalmente, as aplicações de modelação numérica são efetuadas em ambientes convencionais como *workstations* locais, *clusters/grids* e super computadores, cujo estabelecimento, manutenção e operação é complexa. A computação *cloud*, que permite a utilização de máquinas virtuais utilizando computadores e servidores compartilhados e interligados, está a tornar-se uma alternativa atrativa para aplicações com elevadas exigências computacionais, como a modelação de cheias urbanas (Glenis *et al.*, 2013) e a previsão meteorológica (<http://imgi.uibk.ac.at/research/atmospheric-dynamics/projects/raincloud>). As soluções *cloud* permitem o acesso rápido a recursos computacionais sem necessidade de manutenção de recursos próprios, sendo a dimensão dos recursos de computação e armazenamento dinâmica e disponível pelo período de tempo pretendido.

O termo computação *cloud* refere-se às aplicações disponibilizadas como um serviço através da Internet assim como ao próprio *hardware* e *software* dos sistemas (Armbrust *et al.*, 2010). Se o modelo numérico é partilhado por muitos utilizadores, este poderá ser disponibilizado como *Software as a Service* (SaaS), com uma interface externa que permite ao utilizador interagir com os recursos da *cloud* e executar todas as tarefas necessárias para executar o modelo. Para utilizadores avançados que pretendam opções complexas ou utilizar modelos menos comuns, a *cloud* oferece a possibilidade de disponibilizar apenas a infraestrutura como *Infrastructure as a Service* (IaaS).

Atualmente existem ainda poucas aplicações costeiras a utilizarem a *cloud* de forma significativa (e.g. Chandrasekar *et al.*, 2012, Meisinger *et al.*, 2009). Uma das limitações ao seu uso em SPTR prende-se com a imprevisibilidade da performance desta alternativa, quando comparada com ambientes HPC, dada a necessidade diária de repetição do sistema de previsão. Com efeito, os tempos de comunicação da rede podem variar de forma significativa. Para muitas aplicações costeiras a necessidade de simular várias escalas espaciais (do oceano até ao porto) conduz à necessidade de executar os modelos de modo paralelo e distribuído utilizando o paradigma MPI. Os códigos baseados em MPI são otimizados para ambientes HPC (*cluster/grid*), promovendo a comunicação entre recursos e assim tirando partido dos múltiplos nós para alcançar a performance ideal (Costa *et al.*, 2009). Na *cloud*, a replicação deste procedimento pode conduzir a uma redução de performance. Como o uso de federação de *cloud* será necessário (através da integração de recursos *cloud* dispersos), os tempos de comunicação em rede poderão ser significativos e degradar a performance global da simulação.

A modelação de suporte aos SPTR em operação no LNEC tem sido baseada na computação paralela e distribuída de elevado desempenho, habitualmente realizada em *clusters* compostos por múltiplas máquinas independentes interligadas entre si através de redes de comunicação de dados. Dependendo das características das aplicações, o meio de comunicação físico pode ser implementado através de redes Ethernet clássicas ou por redes de baixa latência e alto débito como o *Infiniband*. As redes de baixa latência são particularmente adequadas aos modelos de computação paralelos que exigem um elevado número de processadores em diversas máquinas e comunicação intensa entre as suas instâncias.

Neste sentido, os dois modelos de programação mais usados em aplicações paralelas são o MPI nas aplicações que requerem comunicação entre máquinas independentes, e o OpenMP para programação em máquinas de memória partilhada. Ambos os modelos podem ser combinados para melhor explorar *clusters* de máquinas de memória partilhada.

Estabelecida em 2009, a infraestrutura computacional da iniciativa nacional *grid* tem operado meios de cálculo para suporte à investigação científica em todos os domínios. O seu *cluster* de computação principal encontra-se alojado no LNEC, e suporta serviços de computação sequencial e paralela de elevado desempenho com latência *Ethernet* ou *Infiniband*. Os meios de computação paralela do LNEC encontram-se totalmente integrados neste *cluster* de computação. O *cluster* é ainda complementado por um sistema de armazenamento distribuído de elevado desempenho baseado no sistema de ficheiros Lustre (<http://lustre.org/>).

Em 2014, esta infraestrutura foi selecionada para integrar o roteiro de infraestruturas de interesse estratégico da FCT, designada Infraestrutura Nacional de Computação Distribuída (INCD, www.incd.pt). Neste contexto iniciou-se a experimentação com a computação *cloud*, explorando a possibilidade de disponibilizar serviços do tipo *IaaS* para suporte de aplicações científicas. A computação *cloud* permite ultrapassar algumas das limitações inerentes aos *clusters* de computação tradicionais, proporcionando: 1) maior flexibilidade na alocação de recursos físicos como a memória, CPUs e armazenamento; 2) maior flexibilidade na definição do ambiente computacional (rede, sistema operativo, bibliotecas, e aplicações); 3) melhor isolamento entre utilizadores; e 4) maior autonomia e independência dos utilizadores em relação à infraestrutura.

No entanto muitos destes benefícios baseiam-se na virtualização do processamento, rede e armazenamento, com potencial impacto negativo no desempenho, especialmente em aplicações sensíveis à latência. O estudo efetuado neste artigo foi realizado no contexto do projeto piloto *IaaS* da INCD que visou validar a execução de alguns SPTR do LNEC em ambiente de computação *cloud* e compará-lo com simulações efetuadas noutros ambientes atualmente em modo operacional: *workstations* e *grid cluster*.

5. METODOLOGIA, RECURSOS E DESCRIÇÃO DOS CASOS DE ESTUDO

5.1 Metodologia e recursos computacionais

De modo a assegurar uma análise comparativa adequada e reprodutível entre soluções, foi estabelecido um *benchmark* capaz de produzir métricas comparáveis. O *benchmark* é constituído pela fase mais exigente dos SPTR, a execução dos modelos numéricos, nas várias aplicações. Cada caso de teste foi preparado com os *inputs* necessários para uma simulação diária de previsão. Cada *benchmark* foi executado cinco vezes para produzir valores médios e filtrar variações associadas ao funcionamento dos recursos.

Foram usados dois indicadores de performance: o tempo de execução (minutos) e o ganho de

performance (definido como o rácio entre a performance de duas simulações distintas, medida em passos de cálculo/tempo de computação). Estes dois indicadores permitem analisar em detalhe a capacidade dos vários testes usarem o paradigma MPI e combinarem o poder de computação de vários processadores para atingir melhores performances.

Em termos de recursos, os vários SPTR em estudo executam diariamente em vários ambientes, alguns em ambiente HPC (cluster) outros em *workstations* locais. Esta análise é aqui estendida a um ambiente *cloud*, desenvolvido para testar estas aplicações neste novo tipo de recurso. Descrevem-se abaixo estes vários recursos (quadro 1).

Quadro 1. Características dos recursos computacionais.

Recurso	Processador	Memória	Sistema operativo
Workstation 1	2 x Xeon X5650 2 x 6C @ 2.66Ghz	5 GB	Scientific Linux 5
Workstation 2	2 x Xeon E5645 2 x 6C @ 2.4Ghz	16 GB	Scientific Linux 6
cloud	2 x Xeon E5440 2 x 4C @ 2.83Ghz	24 GB	Scientific Linux 6
Medusa 1	67 x 2 x Opteron 280 67 x 2 x 2C @ 2.4 Ghz	67 x 4 GB (268 GB)	Scientific Linux 6
Medusa 2	4 x Opteron 6220 4 x 8C @ 3Ghz	256 GB	Scientific Linux 6
Medusa 3	2 x Xeon E5-2620 2 x 6C @ 2 Ghz	32 GB	Scientific Linux 6

- **Workstations.** O LNEC detém um *datacenter* que aloja vários servidores e *workstations* utilizados para os vários serviços de tecnologias de informação, nos quais se incluem a operação diária de SPTR e o processamento e publicação na web dos seus resultados. Neste estudo são usados dois recursos: workstation 1 (hydra) e workstation 2 (luna).
- **Medusa cluster/INGRID/INCD.** O *cluster* MEDUSA é um dos recursos integrados no INCD. Os recursos computacionais incluem 3 nós, sendo o primeiro de 2008 e os 2 últimos bastante mais recentes: Medusa 1, Medusa 2 (medusa_it) e Medusa 3 (medusa_intel).
- **Piloto Cloud.** O serviço de computação *cloud* baseia-se em equipamento pré-existente que foi recuperado para implementar um piloto de *IaaS* baseado no *software Openstack* versão Havana (<https://www.openstack.org/>). Este equipamento não possui algumas das funcionalidades mais recentes como a virtualização de I/O. Os servidores foram diretamente ligados via Gigabit Ethernet a um comutador de rede Force10 C300 com 5 microsegundos de latência. A rede foi implementada através do serviço Neutron do Openstack tendo sido virtualizada usando *switches* de rede virtuais Open vSwitch (<http://openvswitch.org/>) interligados por túneis GRE.

O piloto de computação *cloud* do INCD disponibiliza três serviços de armazenamento:

- Armazenamento de imagens de sistemas operativos implementado através do serviço Glance do Openstack.
- Armazenamento de blocos implementado através do serviço Cinder do Openstack. Permite a criação de dispositivos de armazenamento virtuais que podem ser formatados com sistemas de ficheiros convencionais.
- Armazenamento de objetos implementado através do serviço Swift do Openstack. Permite armazenamento distribuído de dados acessíveis através de um serviço web.

5.2 Previsão da circulação para forçamento da evolução de plumas de hidrocarbonetos na Ria de Aveiro

O SPTR de circulação baroclínica e evolução de hidrocarbonetos para a ria de Aveiro tem como principal objetivo otimizar os planos de contingência e apoiar atividades de resposta da Administração do Porto de Aveiro em situações de emergência. Este SPTR fornece diariamente aos gestores costeiros a previsão de 48 horas da circulação hidrodinâmica. Simultaneamente, efetua também simulações de derrames de hidrocarbonetos para seis localizações predefinidas, nos locais onde a ocorrência de um acidente é mais provável, tendo em consideração as atividades portuárias. O conhecimento atempado da evolução de um eventual derrame permite aos gestores uma eficaz coordenação de meios logísticos de forma a mitigar os efeitos nocivos para o ambiente e diversas atividades económicas na Ria de Aveiro. Ambos os modelos utilizam malhas horizontais não-estruturadas permitindo uma representação eficaz de sistemas costeiros com fronteiras complexas e de simulações multi-escala que se podem estender desde o oceano até ao interior de estuários. Além dos resultados do modelo de circulação (e.g. temperatura, salinidade, elevação e velocidades), o SPTR de derrames de hidrocarbonetos também produz diversos indicadores para a gestão de risco, tais como: tempos de exposição ao óleo, evolução das propriedades reológicas do óleo (e.g. densidade, viscosidade, taxa de evaporação, taxa de emulsificação, etc.) e quantidade de óleo retido nas zonas intermareais.

O SPTR é constituído pelo modelo hidrodinâmico SELFE (Zhang e Baptista, 2008) e o modelo de derrames de hidrocarbonetos VOILS (Azevedo *et al.*, 2014). As simulações são realizadas numa malha horizontal não estruturada, com cerca de 23000 nós e 36000 elementos (Figura 3) e uma resolução que varia entre 2 m nos canais mais estreitos no interior da ria e 3 km na zona oceânica. Na vertical, utiliza-se uma malha de coordenadas híbridas SZ com 7 níveis. Os passos de cálculo são de 30 segundos nas simulações hidrodinâmicas e de 900 segundos nas simulações de hidrocarbonetos. Os forçamentos utilizados na aplicação do SPTR para a ria de Aveiro, para níveis, salinidade e temperatura na fronteira oceânica são obtidos do portal MyOcean (www.myocean.eu.org/); os forçamentos de caudais das fronteiras fluviais são obtidos do Sistema Nacional de Informação de Recursos Hídricos (snirh.pt). O modelo VOILS é depois forçado com os resultados do modelo SELFE. A presente análise refere-se apenas às simulações hidrodinâmicas.

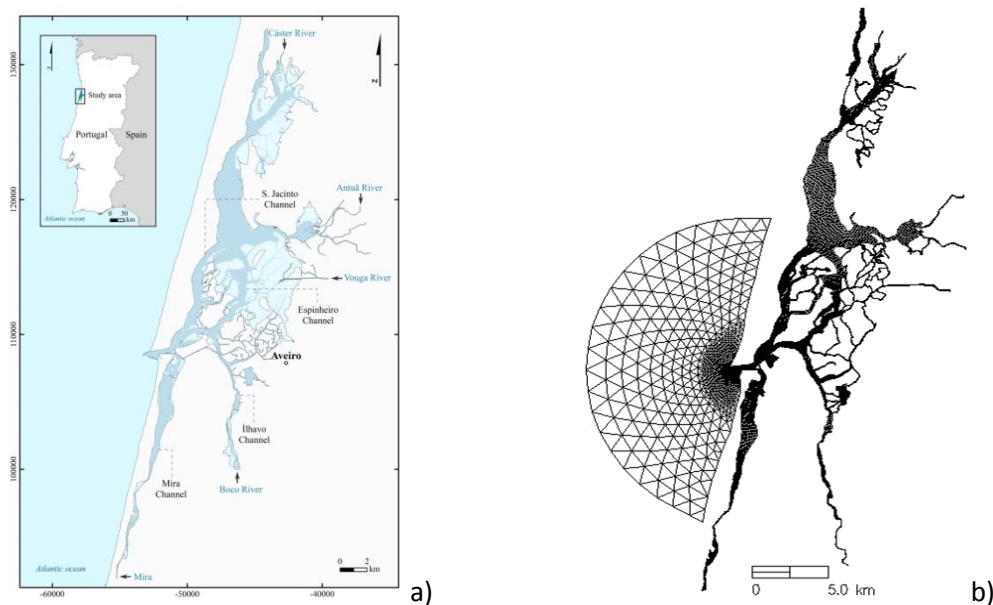


Figura 3: a) Região de estudo e b) malha de cálculo para a Ria de Aveiro e zona marítima adjacente.

5.3 Previsão da circulação baroclínica e da evolução da pluma do Caneiro de Alcântara no estuário do Tejo

O SPTR da circulação baroclínica e de indicadores da contaminação fecal no estuário do Tejo visa a deteção atempada e o alerta de eventos de contaminação neste meio. Este SPTR disponibiliza previsões diárias de níveis, velocidade, salinidade, temperatura e indicadores microbiológicos (coliformes fecais e *Escherichia coli*) (Figura 4a). O sistema piloto foi implementado para as descargas da sub-bacia hidrográfica de Alcântara, a maior da cidade de Lisboa (3200 ha). O caneiro de Alcântara representa o principal ponto de descarga de efluentes urbanos no estuário do Tejo, incluindo os efluentes tratados na estação de tratamento de águas residuais (ETAR) de Alcântara. Este sistema de previsão integra todos os processos físicos e biológicos relevantes, desde a rede de drenagem até ao meio recetor, baseando-se num fluxo de informação entre a rede drenagem, a ETAR e o modelo do estuário do Tejo (Figura 2).

A circulação baroclínica e os indicadores de contaminação fecal no estuário do Tejo são simulados com o modelo acoplado, tridimensional, ECO-SELFE 3.1d (Rodrigues *et al.*, 2009). Rodrigues *et al.* (2013) descreve a implementação, calibração e validação do ECO-SELFE no estuário do Tejo, que foi depois alargada para incluir as descargas relevantes entre Algés-Alcântara-Terreiro do Paço. O domínio é discretizado com uma malha horizontal não estruturada, com cerca de 53000 elementos. A resolução varia entre 1-2 m na zona das descargas e 2 km na zona oceânica. O domínio vertical usa coordenadas híbridas SZ em 20 níveis. O passo de cálculo é de 30 segundos. Em modo operacional, o modelo é forçado por: previsões de níveis, salinidade e temperatura na fronteira oceânica, disponíveis no MyOcean; extrapolações de caudais em tempo real na fronteira do rio Tejo (do SNIRH); e previsões do

modelo de drenagem urbana nas 7 descargas urbanas na zona marginal entre Algés e Terreiro do Paço. O modelo é também forçado por previsões atmosféricas do modelo WRF 9 km (<http://www.windguru.cz>) e do modelo GFS 50km (<http://nomads.ncep.noaa.gov/model>).

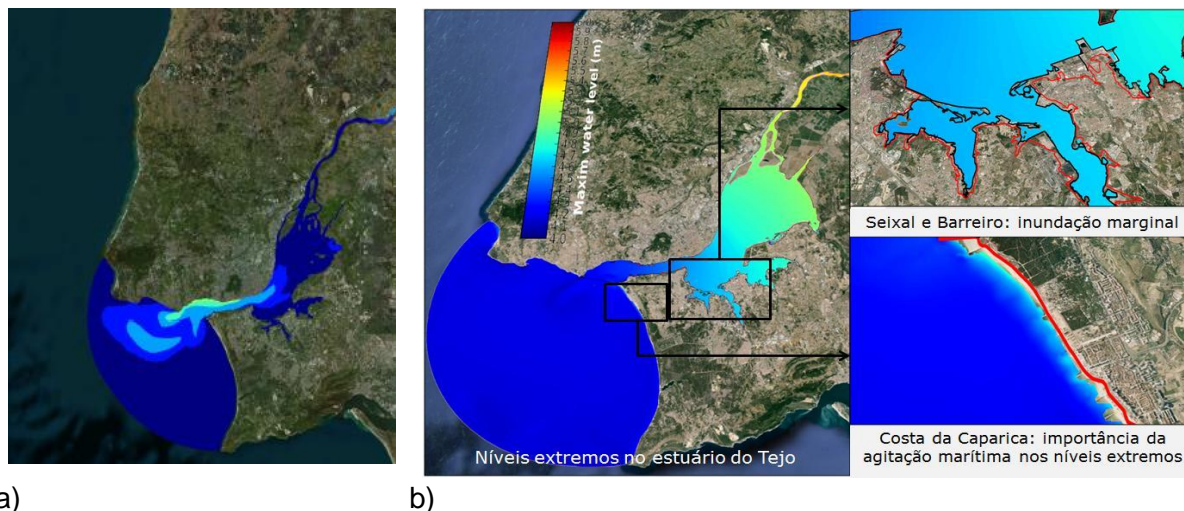


Figura 4 – a) Exemplo de produto do SPTR 3D de circulação e de indicadores de contaminação fecal do estuário do Tejo no WebGIS. b) Exemplos de resultados do modelo de inundação 2D do estuário do Tejo.

5.4 Previsão da circulação acoplada ondas-correntes no estuário do Tejo

O SPTR da circulação acoplada ondas-correntes no estuário do Tejo tem por principal objetivo prever níveis extremos e inundações nas margens do estuário (Figura 4b). É baseado na aplicação do modelo SCHISM-WWM, que simula a hidrodinâmica forçada por marés, ventos e caudais fluviais, a geração e propagação de ondas geradas pelo vento, e a interação entre os dois fenómenos. O modelo é baseado em malhas não-estruturadas e a aplicação ao estuário do Tejo é feita em modo 2D barotrópico. A malha tem cerca de 80000 nós, com uma resolução da ordem da dezena de metros nas zonas inundáveis do estuário. As equações hidrodinâmicas são resolvidas com recurso a fórmulas implícitas, não tendo restrições de estabilidade. Utilizam-se por isso passos de cálculo elevados, de 60 s para a hidrodinâmica e 900 s para a agitação marítima. O modelo é forçado por níveis e agitação marítima provenientes de STPR regionais do LNEC, por previsões de vento e pressão atmosférica fornecidos pela NOAA e pelo caudal fluvial. Simulações a dois dias com 8 processadores demoram cerca de 90 minutos. Uma descrição mais detalhada da implementação e validação deste modelo é apresentada em Fortunato *et al.* (2015).

6. ANÁLISE DE PERFORMANCE

Para a gama de processadores comuns entre os vários ambientes de computação (até 8 cores),

as três alternativas conduzem a resultados similares nos vários testes (Figura 5a), exceto no nó mais antigo (e mais lento) do *cluster* (Medusa 1). Nestas condições, a *cloud* é uma opção viável. No entanto, não foi possível obter o número de processadores ótimo para a aplicação dos vários SPTR em análise dentro da gama comum. A análise revelou ainda que o ganho de performance é ligeiramente superior nos nós do *cluster* que nos restantes recursos (Figura 6a).

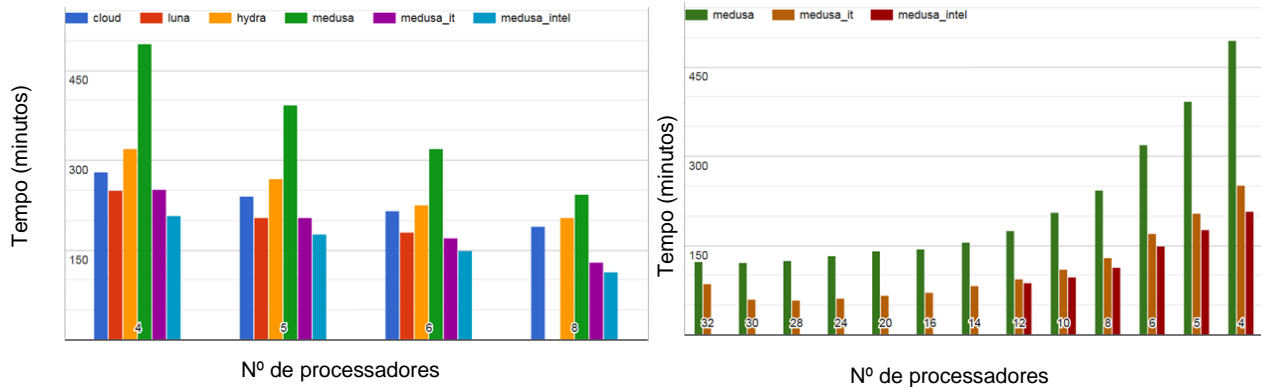


Figura 5. Estuário do Tejo: a) Tempo computacional. b) Procura do tempo computacional ótimo.

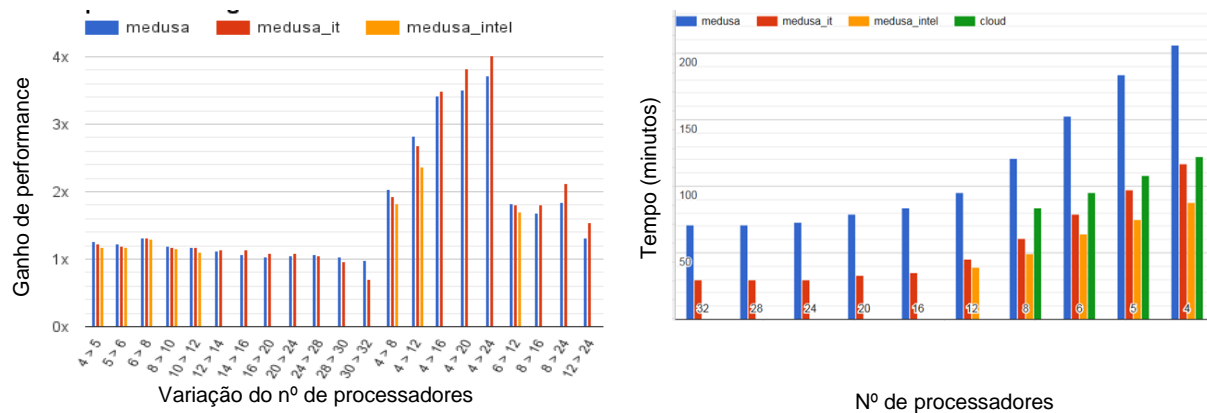


Figura 6. Estuário do Tejo: ganho de performance. Ria de Aveiro: Tempo computacional ótimo.

A procura deste tempo ótimo na Medusa 1 e 2 revelou que este é atingido com cerca de 14/16 cores mas até aos 28/30 a performance continua a aumentar, (apesar do ritmo de aumento ser bem menor) nas duas primeiras aplicações (Figuras 5b e 6b). No entanto este valor depende da configuração numérica do modelo (p.ex. sendo maior para um maior número de nós horizontais ou camadas verticais - Costa *et al.*, 2009). A aplicação de recursos *cloud* para este número ótimo de processadores implicaria a criação de um *cluster* virtual que agruparia várias máquinas virtuais dispersas (em máquinas físicas diferentes), sendo a comunicação entre estas feitas também por uma rede virtual. Dependendo das características do modelo e da latência da comunicação feita entre máquinas virtuais, esta solução de federação *cloud* poderá ter uma degradação de performance e comprometer o seu uso para os SPTR que têm exigências elevadas de garantia de tempo de serviço para a disponibilidade atempada das previsões.

7. CONCLUSÕES

Apresentou-se uma comparação de performance entre *workstations*, *cluster* e *cloud* aplicados a SPTR costeiros operacionais para a dinâmica da Ria de Aveiro e estuário do Tejo baseados em modelos de elevada resolução a operar em paralelo usando MPI. A análise revelou que, para a gama de recursos mais comuns, até 8 cores (limitada no caso dos recursos *cloud* pela máquina física que alberga as máquinas virtuais), as três alternativas conduzem a resultados similares, com exceção para o nó mais antigo do *cluster* por usar hardware manifestamente mais antigo. Nesta gama, verifica-se que os recursos *cloud* são uma opção viável. No entanto, o número ótimo de processadores para estas aplicações excede os recursos da *cloud*, sendo necessário utilizar recursos adicionais através de federação de *cloud*. Neste caso, os tempos de comunicação poderão comprometer a performance da solução, pondo em causa a qualidade do serviço e o acesso atempado às previsões. De futuro pretende-se efetuar esta análise em federação e com equipamentos mais recentes com virtualização de I/O e interfaces de rede do tipo 10GbE com IO-SRV e com *Infiniband*.

AGRADECIMENTOS

Este trabalho foi desenvolvido no âmbito dos projetos MOLINES (PTDC/AAG-MAA/2811/2012), projeto piloto Cloud e INCD, financiados pela Fundação para a Ciência e a Tecnologia (FCT). O 3º autor e a 6ª autora são financiados por bolsas FCT (SFRH/BPD/73089/2010 e SFRH/BPD/87512/2012); o 5º autor é financiado pelo programa doutoral H2Doc.

REFERÊNCIAS BIBLIOGRÁFICAS

- Armbrust, M., A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia, 2010. A view of Cloud computing. *Communications of the ACM*, 53/4, 50–58.
- Azevedo, A. Oliveira, A.B. Fortunato, J. Zhang, A.M. Baptista, 2014. A cross-scale numerical modeling system for management support of oil spill accidents”, *Marine pollution bulletin*, 80/1-2, 132-47.
- Buyya, R., C.S. Yeo, S. Venugopal, 2008. Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities, *Proc. 10th IEEE International Conference on High Performance Computing and Communications*, xii – xv.
- Chandrasekar, K., M. Pathridge, S. Wijeratne, C. Mattocks, and B. Plale, 2012. Middleware alternatives for storm surge predictions in Windows Azure, in “*Proceedings of the 3rd workshop on Scientific Cloud Computing*”, 3-12.

Costa, M, et al. 2009. "Application of Parallel, High-Performance Computing in Coastal Environmental Modeling: Circulation and Ecological Dynamics in the Portuguese Coast", 3rd IBERGRID, 375-386.

Davies, J.R. , V.A. Paramygin, A. Ganguly, R.F. Figueiredo, Y.P. Sheng. 2006. Simulation of storm surge using Grid computing", in "Proceedings of the 9th Estuarine and Coastal Modeling Conference", Spaulding et al (editors), ASCE, 357-374.

Dietrich J.C., C.J.Trahan, M.T.Howard, J.G.Fleming, R.J.Weaver, S.Tanaka, L.Yu, R.A. Luettich Jr., C.N. Dawson, J.J. Westerink, G. Wells, A.Lu, K. Vega, A. Kubach, K.M. Dresback, R.L. Kolar, C. Kaiser, R.R. Twilley (2012). "Surface trajectories of oil transport along the Northern Coastline of the Gulf of Mexico", Continental Shelf Research, 41, 17-47.

Fortunato A.B., Costa R.T., Rogeiro J., Gomes J.L., Li K., Oliveira A., Freire P., Rilo A., Mendes A., Rodrigues M. (2015). Desenvolvimento de um sistema operacional de previsão dos níveis extremos na costa portuguesa, VIII Congresso sobre Planeamento e Gestão das Zonas Costeiras dos Países de Expressão Portuguesa, 15 pp.

Glenis, V., A. S. McGough, V. Kutija, C. Kilsby, S. Woodman, 2013. Flood modelling for cities using Cloud computing", Journal of Cloud Computing: Advances, Systems and Applications, 2/7, 1-14.

Jesus G., Gomes J., Ribeiro N.A., Oliveira A. (2012). "Custom deployment of a Nowcast-forecast information system in coastal regions", Geomundus 2012, 6 pp.

Meisinger, M., C. Farcas, E. Farcas, C. Alexander, M. Arrott, J. de La Beaujardière, P. Hubbard, R. Mendelssohn, R. Signell, 2009. Serving ocean model data on the cloud", OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future, 1-10.

Oliveira, A., G. Jesus, J.L. Gomes, J. Rogeiro, A. Azevedo, M. Rodrigues, A.B. Fortunato, J.M. Dias, L.M. Tomas, E.R. Oliveira, F.L. Alves, S. den Boer. 2014. An interactive WebGIS observatory platform for enhanced support of coastal management, SI. 66, 507-512.

Rodrigues M., et al. 2013. Application of an estuarine and coastal nowcast-forecast information system to the Tagus estuary. Proc. 6th SCACR, 10 pp.

Rodrigues, M., Oliveira, A., Queiroga, H., Fortunato, A., Zhang, Y., (2009). "Three-dimensional modeling of the lower trophic levels in the ria de aveiro (portugal)". Ecological Modelling 220 (9-10), 1274 – 1290.

Wenera, M., J. Schellekens, P. Gijbers, M. van Dijkstra, O. van den Akker, K. Heynert, 2013. The Delft-FEWS flow forecasting system, Environmental Model. & Softw., 40, 65-77.

Zhang, Y., AM. Baptista. 2008. SELFE: A semi-implicit Eulerian-Lagrangian finite-element model for cross scale ocean circulation. Ocean Modelling,21(3-4):71-96.